

(19) World Intellectual Property Organization  
International Bureau(43) International Publication Date  
15 May 2003 (15.05.2003)

PCT

(10) International Publication Number  
WO 03/040325 A2

(51) International Patent Classification <sup>7</sup> :		C12N	60/361,833	5 March 2002 (05.03.2002)	US
			60/362,625	5 March 2002 (05.03.2002)	US
(21) International Application Number:		PCT/US02/35464	60/364,000	13 March 2002 (13.03.2002)	US
			60/364,227	13 March 2002 (13.03.2002)	US
(22) International Filing Date:			60/364,182	13 March 2002 (13.03.2002)	US
5 November 2002 (05.11.2002)			60/364,181	13 March 2002 (13.03.2002)	US
			60/364,197	13 March 2002 (13.03.2002)	US
(25) Filing Language:		English	60/381,621	17 May 2002 (17.05.2002)	US
			60/383,675	28 May 2002 (28.05.2002)	US
(26) Publication Language:		English	60/396,703	17 July 2002 (17.07.2002)	US
			60/401,552	6 August 2002 (06.08.2002)	US
			60/401,594	7 August 2002 (07.08.2002)	US
(30) Priority Data:			60/401,787	7 August 2002 (07.08.2002)	US
60/338,626	5 November 2001 (05.11.2001)	US	60/403,619	15 August 2002 (15.08.2002)	US
60/333,072	6 November 2001 (06.11.2001)	US	60/404,821	20 August 2002 (20.08.2002)	US
60/348,283	9 November 2001 (09.11.2001)	US	60/405,631	23 August 2002 (23.08.2002)	US
60/335,610	15 November 2001 (15.11.2001)	US	60/405,368	23 August 2002 (23.08.2002)	US
60/338,543	16 November 2001 (16.11.2001)	US	60/405,402	23 August 2002 (23.08.2002)	US
60/331,641	20 November 2001 (20.11.2001)	US	60/405,496	23 August 2002 (23.08.2002)	US
60/331,630	20 November 2001 (20.11.2001)	US	60/406,125	26 August 2002 (26.08.2002)	US
60/332,152	21 November 2001 (21.11.2001)	US	10/287,226	4 November 2002 (04.11.2002)	US
60/333,461	27 November 2001 (27.11.2001)	US	(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier applications:		
60/334,027	28 November 2001 (28.11.2001)	US	US	60/338,626 (CIP)	
60/333,912	28 November 2001 (28.11.2001)	US	Filed on	5 November 2001 (05.11.2001)	
60/334,300	29 November 2001 (29.11.2001)	US	US	60/333,072 (CIP)	
60/334,421	30 November 2001 (30.11.2001)	US	Filed on	6 November 2001 (06.11.2001)	
60/334,526	30 November 2001 (30.11.2001)	US	US	60/348,283 (CIP)	
60/336,664	4 December 2001 (04.12.2001)	US	Filed on	9 November 2001 (09.11.2001)	
60/336,576	4 December 2001 (04.12.2001)	US	US	60/335,610 (CIP)	
60/338,314	7 December 2001 (07.12.2001)	US	Filed on	15 November 2001 (15.11.2001)	
60/338,390	7 December 2001 (07.12.2001)	US	US	60/338,543 (CIP)	
60/339,008	10 December 2001 (10.12.2001)	US	Filed on	16 November 2001 (16.11.2001)	
60/339,006	10 December 2001 (10.12.2001)	US	US	60/331,641 (CIP)	
60/339,286	11 December 2001 (11.12.2001)	US	Filed on	20 November 2001 (20.11.2001)	
60/353,288	1 February 2002 (01.02.2002)	US	US	60/331,630 (CIP)	
60/353,280	1 February 2002 (01.02.2002)	US	Filed on	20 November 2001 (20.11.2001)	
60/354,392	4 February 2002 (04.02.2002)	US	US	60/332,152 (CIP)	
60/354,409	4 February 2002 (04.02.2002)	US	Filed on	21 November 2001 (21.11.2001)	
60/354,393	4 February 2002 (04.02.2002)	US	US	60/333,461 (CIP)	
60/360,148	27 February 2002 (27.02.2002)	US	Filed on	27 November 2001 (27.11.2001)	
60/359,944	27 February 2002 (27.02.2002)	US	US	60/334,027 (CIP)	
60/361,790	5 March 2002 (05.03.2002)	US	Filed on	28 November 2001 (28.11.2001)	
60/361,925	5 March 2002 (05.03.2002)	US	US		
60/362,230	5 March 2002 (05.03.2002)	US	Filed on		

[Continued on next page]

(54) Title: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME

(57) Abstract: The present invention provides novel isolated polynucleotides and small molecule target polypeptides encoded by the polynucleotides. Antibodies that immunospecifically bind to a novel small molecule target polypeptide or any derivative, variant, mutant or fragment of that polypeptide, polynucleotide or antibody are disclosed, as are methods in which the small molecule target polypeptide, polynucleotide and antibody are utilized in the detection and treatment of a broad range of pathological states. More specifically, the present invention discloses methods of using recombinantly expressed and/or endogenously expressed proteins in various screening procedures for the purpose of identifying therapeutic antibodies and therapeutic small molecules associated with diseases. The invention further discloses therapeutic, diagnostic and research methods for diagnosis, treatment, and prevention of disorders involving any one of these novel human nucleic acids and proteins.

WO 03/040325 A2



US	60/333,912 (CIP)	US	60/401,594 (CIP)
Filed on	28 November 2001 (28.11.2001)	Filed on	7 August 2002 (07.08.2002)
US	60/334,300 (CIP)	US	60/401,787 (CIP)
Filed on	29 November 2001 (29.11.2001)	Filed on	7 August 2002 (07.08.2002)
US	60/334,421 (CIP)	US	60/403,619 (CIP)
Filed on	30 November 2001 (30.11.2001)	Filed on	15 August 2002 (15.08.2002)
US	60/334,526 (CIP)	US	60/404,821 (CIP)
Filed on	30 November 2001 (30.11.2001)	Filed on	20 August 2002 (20.08.2002)
US	60/336,664 (CIP)	US	60/405,631 (CIP)
Filed on	4 December 2001 (04.12.2001)	Filed on	23 August 2002 (23.08.2002)
US	60/336,576 (CIP)	US	60/405,368 (CIP)
Filed on	4 December 2001 (04.12.2001)	Filed on	23 August 2002 (23.08.2002)
US	60/338,314 (CIP)	US	60/405,402 (CIP)
Filed on	7 December 2001 (07.12.2001)	Filed on	23 August 2002 (23.08.2002)
US	60/338,390 (CIP)		
Filed on	7 December 2001 (07.12.2001)	<b>(71) Applicant (for all designated States except US): CURA-</b>	
US	60/339,008 (CIP)	<b>GEN CORPORATION [US/US]; 555 Long Wharf Drive,</b>	
Filed on	10 December 2001 (10.12.2001)	<b>11th Floor, New Haven, CN 06511 (US).</b>	
US	60/339,006 (CIP)	<b>(72) Inventors; and</b>	
Filed on	10 December 2001 (10.12.2001)	<b>(75) Inventors/Applicants (for US only): AGEE, Michele, L.</b>	
US	60/339,286 (CIP)	<b>[US/US]; 107 Knollwood Road, Wallingford, CT 06492</b>	
Filed on	11 December 2001 (11.12.2001)	<b>(US). ALSOBROOK, John, P., II [US/US]; 60 Lake</b>	
US	60/353,288 (CIP)	<b>Drive, Madison, CT 06443 (US). BERGHS, Constance</b>	
Filed on	1 February 2001 (01.02.2001)	<b>[NL/US]; 459 Orange Street, New Haven, CT 06511 (US).</b>	
US	60/353,280 (CIP)	<b>BOLDOG, Ferenc, L. [HU/US]; 1687 Hartford Turnpike,</b>	
Filed on	1 February 2002 (01.02.2002)	<b>North Haven, CT 06473 (US). BURGESS, Catherine, E.</b>	
US	60/354,392 (CIP)	<b>[US/US]; 90 Carriage Hill Drive, Wethersfield, CT 06109</b>	
Filed on	4 February 2002 (04.02.2002)	<b>(US). CHANT, John, S. [CA/US]; 76 Peddlers Lane,</b>	
US	60/354,409 (CIP)	<b>Branford, CT 06405 (US). CHAUDHURI, Amitabha</b>	
Filed on	4 February 2002 (04.02.2002)	<b>[IN/US]; 99 Harbor Avenue, Madison, CT 06443 (US).</b>	
US	60/354,393 (CIP)	<b>DIPIPPO, Vincent, A. [US/US]; 156 North Atwater</b>	
Filed on	4 February 2002 (04.02.2002)	<b>Street, East Haven, CN 06512 (US). EDINGER, Shlomit,</b>	
US	60/360,148 (CIP)	<b>R. [US/US]; 766 Edgewood Avenue, New Haven, CT</b>	
Filed on	27 February 2002 (27.02.2002)	<b>06515 (US). EISEN, Andrew [US/US]; 12412 St. James</b>	
US	60/359,944 (CIP)	<b>Road, Rockville, MD 20850 (US). ELLERMAN, Karen</b>	
Filed on	27 February 2002 (27.02.2002)	<b>[US/US]; 87 Montoya Drive, Branford, CT 06405 (US).</b>	
US	60/361,790 (CIP)	<b>GANGOLLI, Esha, A. [IN/US]; 31 Strawberry Hill Road,</b>	
Filed on	5 March 2002 (05.03.2002)	<b>Madison, CT 06443 (US). GORMAN, Linda [US/US];</b>	
US	60/361,925 (CIP)	<b>329 Monticello Drive, Branford, CT 06405 (US). GER-</b>	
Filed on	5 March 2002 (05.03.2002)	<b>LACH, Valerie, L. [US/US]; 18 Rock Pasture Road,</b>	
US	60/362,230 (CIP)	<b>Branford, CT 06405 (US). JI, Weizhen [CN/US]; 375</b>	
Filed on	5 March 2002 (05.03.2002)	<b>Old Rock Road, Branford, CT 06405 (US). KEKUDA,</b>	
US	60/361,833 (CIP)	<b>Ramesh [IN/US]; 71 Aiken Street, Unit R3, Norwalk,</b>	
Filed on	5 March 2002 (05.03.2002)	<b>CT 06851 (US). KHRAMTSOV, Nikolai, V. [RU/US];</b>	
US	60/362,625 (CIP)	<b>38 Montoya Circle, Branford, CT 06405 (US). LI, Li</b>	
Filed on	5 March 2002 (05.03.2002)	<b>[CN/US]; 56 Jerimoth Road, Branford, CT 06405 (US).</b>	
US	60/364,000 (CIP)	<b>MALYANKAR, Uriel, M. [IN/US]; 229 Branford Road,</b>	
Filed on	13 March 2002 (13.03.2002)	<b>number 330, Branford, CT 06405 (US). MACDOUGALL,</b>	
US	60/364,227 (CIP)	<b>John, R. [US/US]; 117 Russel Street, Hamden, CT 06517</b>	
Filed on	13 March 2002 (13.03.2002)	<b>(US). MEZES, Peter, S. [CA/US]; 7 Clark's Lane, Old</b>	
US	60/364,182 (CIP)	<b>Lyme, CT 06371 (US). MILLER, Charles, E. [US/US];</b>	
Filed on	13 March 2002 (13.03.2002)	<b>98 Saddle Hill Drive, Guilford, CT 06437 (US). MIL-</b>	
US	60/364,181 (CIP)	<b>LET, Isabelle [FR/US]; 74 Carrington Avenue, Milford,</b>	
Filed on	13 March 2002 (13.03.2002)	<b>CT 06460 (US). OOI, Chean, Eng [US/US]; 14 Flax</b>	
US	60/364,197 (CIP)	<b>Mill Hollow, Branford, CT 06405 (US). ORT, Tatiana</b>	
Filed on	13 March 2002 (13.03.2002)	<b>[RU/US]; 85 Viscount Drive, Apartment 61A, Milford,</b>	
US	60/381,621 (CIP)	<b>CT 06460 (US). PADIGARU, Muralidhara [IN/US]; 71</b>	
Filed on	17 May 2002 (17.05.2002)	<b>Hampton Park, Branford, CT 06405 (US). PATTURA-</b>	
US	60/383,675 (CIP)	<b>JAN, Meera [IN/US]; 45 Harrison Avenue, Apartment 1C,</b>	
Filed on	28 May 2002 (28.05.2002)	<b>Branford, CT 06405 (US). RASTELLI, Luca [IT/US]; 52</b>	
US	60/396,703 (CIP)	<b>Pepperbush Lane, Guilford, CT 06437 (US). RIEGER,</b>	
Filed on	17 July 2002 (17.07.2002)	<b>Daniel, K. [DE/US]; 10A McKinnel Court, Branford,</b>	
US	60/401,552 (CIP)	<b>CT 06405 (US). ROTHENBERG, Mark, E. [US/US]; 2</b>	
Filed on	6 August 2002 (06.08.2002)		

[Continued on next page]



Allen Road, Clinton, CT 06413 (US). **SHENOY, Suresh, G.** [IN/US]; 15 Millwood Drive, Branford, CT 06405 (US). **SPADERNA, Steven, K.** [US/US]; 261 Deerfield Drive, Berlin, CT 06037 (US). **SPYTEK, Kimberly, A.** [US/US]; 28 Court Street, number 1, New Haven, CT 06511 (US). **TAUPIER, Raymond, J., Jr.** [US/US]; 34 Pardee Place Extension, West Haven, CT 06512 (US). **VERNET, Corine, A., M.** [FR/US]; 1739 Foxon Road, Apt. L6, Branford, CT 06471 (US). **ZERHUSEN, Bryan, D.** [US/US]; 337 Monticello Drive, Branford, CT 06405 (US). **ZHONG, Mei** [CA/US]; 152 Brushy Plain Road, Branford, CT 06405 (US).

(74) **Agent: ELRIFI, Ivor, R.**; Mintz, Levin, Cohn, Ferris, Glovsky and Popeo, P.C., One Financial Center, Boston, MA 02111 (US).

(81) **Designated States (national):** AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC,

LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) **Designated States (regional):** ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

**Published:**

— without international search report and to be republished upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

## NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME

### FIELD OF THE INVENTION

The present invention relates to novel polypeptides that are targets of small  
5 molecule drugs and that have properties related to stimulation of biochemical or  
physiological responses in a cell, a tissue, an organ or an organism. More particularly, the  
novel polypeptides are gene products of novel genes, or are specified biologically active  
fragments or derivatives thereof. Methods of use encompass diagnostic and prognostic  
assay procedures as well as methods of treating diverse pathological conditions.  
10

U.S.S.N. 60/406125, filed August 26, 2002; U.S.S.N. 60/338543, filed November 16, 2001; U.S.S.N. 60/339286, filed December 11, 2001; U.S.S.N. 60/336576, filed December 4, 2001; U.S.S.N. 60/333912, filed November 28, 2001; each of which is incorporated herein by reference in its entirety.

5

## FIELD OF THE INVENTION

The present invention relates to novel polypeptides that are targets of small molecule drugs and that have properties related to stimulation of biochemical or physiological responses in a cell, a tissue, an organ or an organism. More particularly, the novel polypeptides are gene products of novel genes, or are specified biologically active fragments or derivatives thereof. Methods of use encompass diagnostic and prognostic assay procedures as well as methods of treating diverse pathological conditions.

10

## BACKGROUND

Eukaryotic cells are characterized by biochemical and physiological processes which under normal conditions are exquisitely balanced to achieve the preservation and propagation of the cells. When such cells are components of multicellular organisms such as vertebrates, or more particularly organisms such as mammals, the regulation of the biochemical and physiological processes involves intricate signaling pathways. Frequently, such signaling pathways involve extracellular signaling proteins, cellular receptors that bind the signaling proteins and signal transducing components located within the cells.

Signaling proteins may be classified as endocrine effectors, paracrine effectors or autocrine effectors. Endocrine effectors are signaling molecules secreted by a given organ into the circulatory system, which are then transported to a distant target organ or tissue. The target cells include the receptors for the endocrine effector, and when the endocrine effector binds, a signaling cascade is induced. Paracrine effectors involve secreting cells and receptor cells in close proximity to each other, for example two different classes of cells in the same tissue or organ. One class of cells secretes the paracrine effector, which then reaches the second class of cells, for example by diffusion through the extracellular fluid. The second class of cells contains the receptors for the paracrine effector; binding of the effector results in induction of the signaling cascade that elicits the corresponding biochemical or physiological effect. Autocrine effectors are highly analogous to paracrine effectors, except that the same cell type that secretes the autocrine effector also contains the receptor. Thus the autocrine effector binds to receptors on the same cell, or on identical neighboring cells. The binding process then elicits the characteristic biochemical or physiological effect.

Signaling processes may elicit a variety of effects on cells and tissues including by way of nonlimiting example induction of cell or tissue proliferation, suppression of growth or proliferation, induction of differentiation or maturation of a cell or tissue, and suppression of differentiation or maturation of a cell or tissue.

Many pathological conditions involve dysregulation of expression of important effector proteins. In certain classes of pathologies the dysregulation is manifested as diminished or suppressed level of synthesis and secretion of protein effectors. In other classes of pathologies the dysregulation is manifested as increased or up-regulated level of synthesis and secretion of protein effectors. In a clinical setting a subject may be suspected of suffering from a condition brought on by altered or mis-regulated levels of a protein

effector of interest. Therefore there is a need to assay for the level of the protein effector of interest in a biological sample from such a subject, and to compare the level with that characteristic of a nonpathological condition. There also is a need to provide the protein effector as a product of manufacture. Administration of the effector to a subject in need thereof is useful in treatment of the pathological condition. Accordingly, there is a need for a method of treatment of a pathological condition brought on by a diminished or suppressed levels of the protein effector of interest. In addition, there is a need for a method of treatment of a pathological condition brought on by a increased or up-regulated levels of the protein effector of interest.

Small molecule targets have been implicated in various disease states or pathologies. These targets may be proteins, and particularly enzymatic proteins, which are acted upon by small molecule drugs for the purpose of altering target function and achieving a desired result. Cellular, animal and clinical studies can be performed to elucidate the genetic contribution to the etiology and pathogenesis of conditions in which small molecule targets are implicated in a variety of physiologic, pharmacologic or native states. These studies utilize the core technologies at CuraGen Corporation to look at differential gene expression, protein-protein interactions, large-scale sequencing of expressed genes and the association of genetic variations such as, but not limited to, single nucleotide polymorphisms (SNPs) or splice variants in and between biological samples from experimental and control groups. The goal of such studies is to identify potential avenues for therapeutic intervention in order to prevent, treat the consequences or cure the conditions.

In order to treat diseases, pathologies and other abnormal states or conditions in which a mammalian organism has been diagnosed as being, or as being at risk for becoming, other than in a normal state or condition, it is important to identify new therapeutic agents. Such a procedure includes at least the steps of identifying a target component within an affected tissue or organ, and identifying a candidate therapeutic agent that modulates the functional attributes of the target. The target component may be any biological macromolecule implicated in the disease or pathology. Commonly the target is a polypeptide or protein with specific functional attributes. Other classes of macromolecule may be a nucleic acid, a polysaccharide, a lipid such as a complex lipid or a glycolipid; in addition a target may be a sub-cellular structure or extra-cellular structure that is comprised of more than one of these classes of macromolecule. Once such a target has been

identified, it may be employed in a screening assay in order to identify favorable candidate therapeutic agents from among a large population of substances or compounds.

In many cases the objective of such screening assays is to identify small molecule candidates; this is commonly approached by the use of combinatorial methodologies to develop the population of substances to be tested. The implementation of high throughput screening methodologies is advantageous when working with large, combinatorial libraries of compounds.

### SUMMARY OF THE INVENTION

The invention includes nucleic acid sequences and the novel polypeptides they encode... The novel nucleic acids and polypeptides are referred to herein as NOVX, or NOV1, NOV2, NOV3, *etc.*, nucleic acids and polypeptides. These nucleic acids and polypeptides, as well as derivatives, homologs, analogs and fragments thereof, will hereinafter be collectively designated as "NOVX" nucleic acid, which represents the nucleotide sequence selected from the group consisting of SEQ ID NO: 2n-1, wherein n is an integer between 1 and 226, or polypeptide sequences, which represents the group consisting of SEQ ID NO: 2n, wherein n is an integer between 1 and 226.

In one aspect, the invention provides an isolated polypeptide comprising a mature form of a NOVX amino acid. One example is a variant of a mature form of a NOVX amino acid sequence, wherein any amino acid in the mature form is changed to a different amino acid, provided that no more than 15% of the amino acid residues in the sequence of the mature form are so changed. The amino acid can be, for example, a NOVX amino acid sequence or a variant of a NOVX amino acid sequence, wherein any amino acid specified in the chosen sequence is changed to a different amino acid, provided that no more than 15% of the amino acid residues in the sequence are so changed. The invention also includes fragments of any of these. In another aspect, the invention also includes an isolated nucleic acid that encodes a NOVX polypeptide, or a fragment, homolog, analog or derivative thereof.

Also included in the invention is a NOVX polypeptide that is a naturally occurring allelic variant of a NOVX sequence. In one embodiment, the allelic variant includes an amino acid sequence that is the translation of a nucleic acid sequence differing by a single nucleotide from a NOVX nucleic acid sequence. In another embodiment, the NOVX polypeptide is a variant polypeptide described therein, wherein any amino acid specified in the chosen sequence is changed to provide a conservative substitution. In one embodiment,

the invention discloses a method for determining the presence or amount of the NOVX polypeptide in a sample. The method involves the steps of: providing a sample; introducing the sample to an antibody that binds immunospecifically to the polypeptide; and determining the presence or amount of antibody bound to the NOVX polypeptide, thereby determining the presence or amount of the NOVX polypeptide in the sample. In another embodiment, the invention provides a method for determining the presence of or predisposition to a disease associated with altered levels of a NOVX polypeptide in a mammalian subject. This method involves the steps of: measuring the level of expression of the polypeptide in a sample from the first mammalian subject; and comparing the amount of the polypeptide in the sample of the first step to the amount of the polypeptide present in a control sample from a second mammalian subject known not to have, or not to be predisposed to, the disease, wherein an alteration in the expression level of the polypeptide in the first subject as compared to the control sample indicates the presence of or predisposition to the disease.

In a further embodiment, the invention includes a method of identifying an agent that binds to a NOVX polypeptide. This method involves the steps of: introducing the polypeptide to the agent; and determining whether the agent binds to the polypeptide. In various embodiments, the agent is a cellular receptor or a downstream effector.

In another aspect, the invention provides a method for identifying a potential therapeutic agent for use in treatment of a pathology, wherein the pathology is related to aberrant expression or aberrant physiological interactions of a NOVX polypeptide. The method involves the steps of: providing a cell expressing the NOVX polypeptide and having a property or function ascribable to the polypeptide; contacting the cell with a composition comprising a candidate substance; and determining whether the substance alters the property or function ascribable to the polypeptide; whereby, if an alteration observed in the presence of the substance is not observed when the cell is contacted with a composition devoid of the substance, the substance is identified as a potential therapeutic agent. In another aspect, the invention describes a method for screening for a modulator of activity or of latency or predisposition to a pathology associated with the NOVX polypeptide. This method involves the following steps: administering a test compound to a test animal at increased risk for a pathology associated with the NOVX polypeptide, wherein the test animal recombinantly expresses the NOVX polypeptide. This method involves the steps of measuring the activity of the NOVX polypeptide in the test animal

after administering the compound of step; and comparing the activity of the protein in the test animal with the activity of the NOVX polypeptide in a control animal not administered the polypeptide, wherein a change in the activity of the NOVX polypeptide in the test animal relative to the control animal indicates the test compound is a modulator of latency of, or predisposition to, a pathology associated with the NOVX polypeptide. In one embodiment, the test animal is a recombinant test animal that expresses a test protein transgene or expresses the transgene under the control of a promoter at an increased level relative to a wild-type test animal, and wherein the promoter is not the native gene promoter of the transgene. In another aspect, the invention includes a method for modulating the activity of the NOVX polypeptide, the method comprising introducing a cell sample expressing the NOVX polypeptide with a compound that binds to the polypeptide in an amount sufficient to modulate the activity of the polypeptide.

The invention also includes an isolated nucleic acid that encodes a NOVX polypeptide, or a fragment, homolog, analog or derivative thereof. In a preferred embodiment, the nucleic acid molecule comprises the nucleotide sequence of a naturally occurring allelic nucleic acid variant. In another embodiment, the nucleic acid encodes a variant polypeptide, wherein the variant polypeptide has the polypeptide sequence of a naturally occurring polypeptide variant. In another embodiment, the nucleic acid molecule differs by a single nucleotide from a NOVX nucleic acid sequence. In one embodiment, the NOVX nucleic acid molecule hybridizes under stringent conditions to the nucleotide sequence selected from the group consisting of SEQ ID NO: 2n-1, wherein n is an integer between 1 and 226, or a complement of the nucleotide sequence. In another aspect, the invention provides a vector or a cell expressing a NOVX nucleotide sequence.

In one embodiment, the invention discloses a method for modulating the activity of a NOVX polypeptide. The method includes the steps of: introducing a cell sample expressing the NOVX polypeptide with a compound that binds to the polypeptide in an amount sufficient to modulate the activity of the polypeptide. In another embodiment, the invention includes an isolated NOVX nucleic acid molecule comprising a nucleic acid sequence encoding a polypeptide comprising a NOVX amino acid sequence or a variant of a mature form of the NOVX amino acid sequence, wherein any amino acid in the mature form of the chosen sequence is changed to a different amino acid, provided that no more than 15% of the amino acid residues in the sequence of the mature form are so changed. In another embodiment, the invention includes an amino acid sequence that is a variant of the

NOVX amino acid sequence, in which any amino acid specified in the chosen sequence is changed to a different amino acid, provided that no more than 15% of the amino acid residues in the sequence are so changed.

- In one embodiment, the invention discloses a NOVX nucleic acid fragment
- 5 encoding at least a portion of a NOVX polypeptide or any variant of the polypeptide, wherein any amino acid of the chosen sequence is changed to a different amino acid, provided that no more than 10% of the amino acid residues in the sequence are so changed. In another embodiment, the invention includes the complement of any of the NOVX nucleic acid molecules or a naturally occurring allelic nucleic acid variant. In another
- 10 embodiment, the invention discloses a NOVX nucleic acid molecule that encodes a variant polypeptide, wherein the variant polypeptide has the polypeptide sequence of a naturally occurring polypeptide variant. In another embodiment, the invention discloses a NOVX nucleic acid, wherein the nucleic acid molecule differs by a single nucleotide from a NOVX nucleic acid sequence.
- 15 In another aspect, the invention includes a NOVX nucleic acid, wherein one or more nucleotides in the NOVX nucleotide sequence is changed to a different nucleotide provided that no more than 15% of the nucleotides are so changed. In one embodiment, the invention discloses a nucleic acid fragment of the NOVX nucleotide sequence and a nucleic acid fragment wherein one or more nucleotides in the NOVX nucleotide sequence
- 20 is changed from that selected from the group consisting of the chosen sequence to a different nucleotide provided that no more than 15% of the nucleotides are so changed. In another embodiment, the invention includes a nucleic acid molecule wherein the nucleic acid molecule hybridizes under stringent conditions to a NOVX nucleotide sequence or a complement of the NOVX nucleotide sequence. In one embodiment, the invention
- 25 includes a nucleic acid molecule, wherein the sequence is changed such that no more than 15% of the nucleotides in the coding sequence differ from the NOVX nucleotide sequence or a fragment thereof.

- In a further aspect, the invention includes a method for determining the presence or amount of the NOVX nucleic acid in a sample. The method involves the steps of:
- 30 providing the sample; introducing the sample to a probe that binds to the nucleic acid molecule; and determining the presence or amount of the probe bound to the NOVX nucleic acid molecule, thereby determining the presence or amount of the NOVX nucleic

acid molecule in the sample. In one embodiment, the presence or amount of the nucleic acid molecule is used as a marker for cell or tissue type.

In another aspect, the invention discloses a method for determining the presence of or predisposition to a disease associated with altered levels of the NOVX nucleic acid molecule of in a first mammalian subject. The method involves the steps of: measuring the amount of NOVX nucleic acid in a sample from the first mammalian subject; and comparing the amount of the nucleic acid in the sample of step (a) to the amount of NOVX nucleic acid present in a control sample from a second mammalian subject known not to have or not be predisposed to, the disease; wherein an alteration in the level of the nucleic acid in the first subject as compared to the control sample indicates the presence of or predisposition to the disease.

Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Although methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, suitable methods and materials are described below. All publications, patent applications, patents, and other references mentioned herein are incorporated by reference in their entirety. In the case of conflict, the present specification, including definitions, will control. In addition, the materials, methods, and examples are illustrative only and not intended to be limiting.

Other features and advantages of the invention will be apparent from the following detailed description and claims.

#### DETAILED DESCRIPTION OF THE INVENTION

The present invention provides novel nucleotides and polypeptides encoded thereby. Included in the invention are the novel nucleic acid sequences, their encoded polypeptides, antibodies, and other related compounds. The sequences are collectively referred to herein as "NOVX nucleic acids" or "NOVX polynucleotides" and the corresponding encoded polypeptides are referred to as "NOVX polypeptides" or "NOVX proteins." Unless indicated otherwise, "NOVX" is meant to refer to any of the novel sequences disclosed herein. Table A provides a summary of the NOVX nucleic acids and their encoded polypeptides.

TABLE A. Sequences and Corresponding SEQ ID Numbers

NOVX Assignment	Internal Identification	SEQ ID NO (nucleic acid)	SEQ ID NO (amino acid)	Homology
1a	CG101683-01	1	2	Mitogen-activated protein kinase kinase kinase 8
1b	248490507	3	4	Mitogen-activated protein kinase kinase kinase 8
1c	253174293	5	6	Mitogen-activated protein kinase kinase kinase 8
1d	248490584	7	8	Mitogen-activated protein kinase kinase kinase 8
1e	258054391	9	10	Mitogen-activated protein kinase kinase kinase 8
1f	248494549	11	12	Mitogen-activated protein kinase kinase kinase 8
1g	259741837	13	14	Mitogen-activated protein kinase kinase kinase 8
1h	260480803	15	16	Mitogen-activated protein kinase kinase kinase 8
1i	209983329	17	18	Mitogen-activated protein kinase kinase kinase 8
1j	212779055	19	20	Mitogen-activated protein kinase kinase kinase 8
1k	212779063	21	22	Mitogen-activated protein kinase kinase kinase 8
1l	CG101683-02	23	24	Mitogen-activated protein kinase kinase kinase 8
1m	CG101683-03	25	26	Mitogen-activated protein kinase kinase kinase 8
1n	CG101683-04	27	28	Mitogen-activated protein kinase kinase kinase 8
1o	CG101683-05	29	30	Mitogen-activated protein kinase kinase kinase 8
1p	CG101683-06	31	32	Mitogen-activated protein kinase kinase kinase 8
1q	CG101683-07	33	34	Mitogen-activated protein kinase kinase kinase 8
1r	CG101683-08	35	36	Mitogen-activated protein kinase kinase kinase 8
2a	CG101996-01	37	38	Phosphorylase B kinase gamma catalytic chain, skeletal muscle isoform
2b	CG101996-04	39	40	Phosphorylase B kinase gamma catalytic chain, skeletal muscle isoform
2c	CG101996-02	41	42	Phosphorylase B kinase gamma catalytic chain, skeletal muscle isoform
2d	245245680	43	44	Phosphorylase B kinase gamma catalytic chain, skeletal muscle isoform
2e	245245707	45	46	Phosphorylase B kinase gamma catalytic chain, skeletal muscle isoform
2f	248494552	47	48	Phosphorylase B kinase gamma catalytic chain, skeletal muscle isoform
2g	242435676	49	50	Phosphorylase B kinase gamma catalytic chain,

				skeletal muscle isoform
2h	254868664	51	52	Phosphorylase B kinase gamma catalytic chain, skeletal muscle isoform
2i	249122191	53	54	Phosphorylase B kinase gamma catalytic chain, skeletal muscle isoform
2j	249122234	55	56	Phosphorylase B kinase gamma catalytic chain, skeletal muscle isoform
2k	CG101996-03	57	58	Phosphorylase B kinase gamma catalytic chain, skeletal muscle isoform
2l	CG101996-05	59	60	Phosphorylase B kinase gamma catalytic chain, skeletal muscle isoform
2m	CG101996-06	61	62	Phosphorylase B kinase gamma catalytic chain, skeletal muscle isoform
2n	CG101996-07	63	64	Phosphorylase B kinase gamma catalytic chain, skeletal muscle isoform
2o	CG101996-08	65	66	Phosphorylase B kinase gamma catalytic chain, skeletal muscle isoform
2p	CG101996-09	67	68	Phosphorylase B kinase gamma catalytic chain, skeletal muscle isoform
3a	CG102822-01	69	70	glutamate--ammonia ligase
3b	CG102822-03	71	72	glutamate--ammonia ligase
3c	CG102822-03	73	74	glutamate--ammonia ligase
3d	CG102822-04	75	76	glutamate--ammonia ligase
4a	CG103241-01	77	78	Beta-1,4- galactosyltransferase 2
4b	CG103241-02	79	80	Beta-1,4- galactosyltransferase 2
4c	CG103241-03	81	82	Beta-1,4- galactosyltransferase 2
5a	CG106249-01	83	84	KIAA1590 protein
5b	CG106249-02	85	86	KIAA1590 protein
6a	CG106824-01	87	88	Tryptase beta-1 precursor
6b	CG106824-04	89	90	Tryptase beta-1 precursor
6c	CG106824-02	91	92	Tryptase beta-1 precursor
6d	CG106824-03	93	94	Tryptase beta-1 precursor
7a	CG114327-01	95	96	Similar to hypothetical protein FLJ23469
7b	CG114327-02	97	98	Similar to hypothetical protein FLJ23469
8a	CG119418-01	99	100	Farnesyl-diphosphate farnesyltransferase
9a	CG120359-01	101	102	Acetyl-coenzyme A synthetase, cytoplasmic

9b	277685717	103	104	Acetyl-coenzyme A synthetase, cytoplasmic
9c	277686882	105	106	Acetyl-coenzyme A synthetase, cytoplasmic
9d	CG120359-02	107	108	Acetyl-coenzyme A synthetase, cytoplasmic
10a	CG124907-01	109	110	Ornithine decarboxylase
10b	CG124907-01	111	112	Ornithine decarboxylase
10c	254048022	113	114	Ornithine decarboxylase
10d	258252457	115	116	Ornithine decarboxylase
10e	258280014	117	118	Ornithine decarboxylase
10f	258330318	119	120	Ornithine decarboxylase
10g	258330346	121	122	Ornithine decarboxylase
10h	258330472	123	124	Ornithine decarboxylase
10i	258330611	125	126	Ornithine decarboxylase
10j	260481330	127	128	Ornithine decarboxylase
10k	CG124907-02	129	130	Ornithine decarboxylase
10l	CG124907-03	131	132	Ornithine decarboxylase
10m	CG124907-04	133	134	Ornithine decarboxylase
10n	CG124907-05	135	136	Ornithine decarboxylase
10o	CG124907-06	137	138	Ornithine decarboxylase
11a	CG128347-01	139	140	Hypothetical 96.7 kDa protein
11b	CG128347-02	141	142	Hypothetical 96.7 kDa protein
12a	CG135823-01	143	144	Tyrosine aminotransferase
12b	CG135823-02	145	146	Tyrosine aminotransferase
12c	233048273	147	148	Tyrosine aminotransferase
12d	233048286	149	150	Tyrosine aminotransferase
12e	248490358	151	152	Tyrosine aminotransferase
12f	254868693	153	154	Tyrosine aminotransferase
12g	255667122	155	156	Tyrosine aminotransferase
12h	258252417	157	158	Tyrosine aminotransferase
12i	259741773	159	160	Tyrosine aminotransferase
12j	260480043	161	162	Tyrosine aminotransferase
12k	CG135823-03	163	164	Tyrosine aminotransferase
12l	CG135823-04	165	166	Tyrosine aminotransferase
13a	CG140122-01	167	168	Polyamine oxidase isoform-1 - Homo sapiens

13b	246864043	169	170	Polyamine oxidase isoform-1 - Homo sapiens
13c	246864086	171	172	Polyamine oxidase isoform-1 - Homo sapiens
13d	258280083	173	174	Polyamine oxidase isoform-1 - Homo sapiens
13e	258280066	175	176	Polyamine oxidase isoform-1 - Homo sapiens
13f	258329988	177	178	Polyamine oxidase isoform-1 - Homo sapiens
13g	254047897	179	180	Polyamine oxidase isoform-1 - Homo sapiens
13h	258329988	181	182	Polyamine oxidase isoform-1 - Homo sapiens
13i	258280066	183	184	Polyamine oxidase isoform-1 - Homo sapiens
13j	258280083	185	186	Polyamine oxidase isoform-1 - Homo sapiens
13k	CG140122-02	187	188	Polyamine oxidase isoform-1 - Homo sapiens
13l	CG140122-03	189	190	Polyamine oxidase isoform-1 - Homo sapiens
13m	CG140122-04	191	192	Polyamine oxidase isoform-1 - Homo sapiens
13n	CG140122-05	193	194	Polyamine oxidase isoform-1 - Homo sapiens
13o	CG140122-06	195	196	Polyamine oxidase isoform-1 - Homo sapiens
13p	CG140122-07	197	198	Polyamine oxidase isoform-1 - Homo sapiens
13q	CG140122-08	199	200	Polyamine oxidase isoform-1 - Homo sapiens
14a	CG140316-01	201	202	NADP-dependent malic enzyme
14b	CG140316-01	203	204	NADP-dependent malic enzyme
14c	254047949	205	206	NADP-dependent malic enzyme
14d	258280122	207	208	NADP-dependent malic enzyme
14e	258330149	209	210	NADP-dependent malic enzyme
14f	258330422	211	212	NADP-dependent malic enzyme
14g	258330562	213	214	NADP-dependent malic enzyme
14h	258330639	215	216	NADP-dependent malic enzyme
14i	259357792	217	218	NADP-dependent malic enzyme
14j	CG140316-02	219	220	NADP-dependent malic enzyme
14k	CG140316-03	221	222	NADP-dependent malic enzyme
14l	CG140316-04	223	224	NADP-dependent malic enzyme
15a	CG142427-01	225	226	ATP-citrate (pro-S)-lyase
15b	CG142427-01	227	228	ATP-citrate (pro-S)-lyase

15c	CG142427-04	229	230	ATP-citrate (pro-S-)-lyase
15d	CG142427-02	231	232	ATP-citrate (pro-S-)-lyase
15e	CG142427-03	233	234	ATP-citrate (pro-S-)-lyase
15f	256388552	235	236	ATP-citrate (pro-S-)-lyase
15g	256420210	237	238	ATP-citrate (pro-S-)-lyase
15h	256202925	239	240	ATP-citrate (pro-S-)-lyase
15i	259856081	241	242	ATP-citrate (pro-S-)-lyase
15j	256388552	243	244	ATP-citrate (pro-S-)-lyase
15k	256420210	245	246	ATP-citrate (pro-S-)-lyase
15l	256202925	247	248	ATP-citrate (pro-S-)-lyase
15m	296463359	249	250	ATP-citrate (pro-S-)-lyase
15n	263470992	251	252	ATP-citrate (pro-S-)-lyase
15o	CG142427-05	253	254	ATP-citrate (pro-S-)-lyase
16a	CG142631-01	255	256	L-serine dehydratase
16b	CG142631-01	257	258	L-serine dehydratase
16c	248494617	259	260	L-serine dehydratase
16d	228832711	261	262	L-serine dehydratase
16e	256420310	263	264	L-serine dehydratase
16f	249117058	265	266	L-serine dehydratase
16g	252790334	267	268	L-serine dehydratase
16h	254869149	269	270	L-serine dehydratase
16i	CG142631-02	271	272	L-serine dehydratase
16j	CG142631-03	273	274	L-serine dehydratase
16k	CG142631-04	275	276	L-serine dehydratase
17a	CG151359-01	277	278	L-lactate dehydrogenase A-like
18a	CG152227-01	279	280	Similar to 3-hydroxyisobutyryl-coenzyme A hydrolase
18b	CG152227-02	281	282	Similar to 3-hydroxyisobutyryl-coenzyme A hydrolase
19a	CG152392-01	283	284	Hypothetical 68.5 kDa protein
20a	CG152453-01	285	286	Beta-1,4-galactosyltransferase 6
20b	CG152453-03	287	288	Beta-1,4-galactosyltransferase 6
20c	CG152453-02	289	290	Beta-1,4-galactosyltransferase 6
21a	CG152547-01	291	292	Hypothetical 26.3 kDa protein
22a	CG152646-01	293	294	Hypothetical 57.5 kDa protein
23a	CG152959-01	295	296	CAAX prenyl protease 2

23b	CG152959-02	297	298	CAAX prenyl protease 2
24a	CG153033-01	299	300	Vesicular glutamate transporter 3 - Homo sapiens
25a	CG153818-01	301	302	CDNA FLJ37300 fis, clone BRAMY2015782, moderately similar to KINESIN-LIKE PROTEIN
26a	CG154435-01	303	304	Dynein beta chain, ciliary
27a	CG154465-01	305	306	Similar to hypothetical protein DKFZp434G2226 -
28a	CG154492-01	307	308	High-affinity cGMP-specific 3',5'-cyclic phosphodiesterase 9A
28b	CG154492-02	309	310	High-affinity cGMP-specific 3',5'-cyclic phosphodiesterase 9A
29a	CG154509-01	311	312	Cytoplasmic dynein heavy chain
30a	CG155595-01	313	314	Hypothetical 98.5 kDa protein
31a	CG155962-01	315	316	Kinesin-like protein KIF1B (Klp)
32a	CG157477-01	317	318	Myosin I
33a	CG157486-01	319	320	EphA2
34a	CG157505-01	321	322	KIAA1300 protein
35a	CG157629-01	323	324	Serine/threonine protein phosphatase with EF-hands-1
35b	CG157629-01	325	326	Serine/threonine protein phosphatase with EF-hands-1
36a	CG157704-01	327	328	Probable mitotic centromere associated kinesin - Leishmania major
37a	CG158218-01	329	330	Kinesin-related protein 3A
38a	CG158513-01	331	332	Prostatic acid phosphatase precursor
38b	CG158513-02	333	334	Prostatic acid phosphatase precursor
39a	CG158583-01	335	336	Synaptic vesicle amine transporter (Monoamine transporter) (Vesicular amine transporter 2) (VAT2)
39b	CG158583-02	337	338	Synaptic vesicle amine transporter (Monoamine transporter) (Vesicular amine transporter 2) (VAT2)
39c	CG158583-04	339	340	Synaptic vesicle amine transporter (Monoamine transporter) (Vesicular

				amine transporter 2) (VAT2)
39d	CG158583-05	341	342	Synaptic vesicle amine transporter (Monoamine transporter) (Vesicular amine transporter 2) (VAT2)
39e	CG158583-03	343	345	Synaptic vesicle amine transporter (Monoamine transporter) (Vesicular amine transporter 2) (VAT2)
40a	CG158964-01	346	347	PHOSPHATIDIC acid phosphatase 2A
40b	CG158964-02	348	349	PHOSPHATIDIC acid phosphatase 2A
41a	CG159084-01	349	350	Glutamate decarboxylase 67
42a	CG159130-01	351	352	Hyperpolarization- activated cation channel, HAC2
43a	CG159178-01	353	354	Carbonic anhydrase VI precursor (EC 4.2.1.1) (Carbonate dehydratase VI) (CA-VI) (Secreted carbonic anhydrase) (Salivary carbonic anhydrase)
43b	CG159178-02	355	356	Carbonic anhydrase VI precursor (EC 4.2.1.1) (Carbonate dehydratase VI) (CA-VI) (Secreted carbonic anhydrase) (Salivary carbonic anhydrase)
44a	CG160131-01	357	358	Glycerol kinase (EC 2.7.1.30) (ATP:glycerol 3-phosphotransferase) (Glycerokinase) (GK)
44b	CG160131-04	359	360	Glycerol kinase (EC 2.7.1.30) (ATP:glycerol 3-phosphotransferase) (Glycerokinase) (GK)
44c	CG160131-02	361	362	Glycerol kinase (EC 2.7.1.30) (ATP:glycerol 3-phosphotransferase) (Glycerokinase) (GK)
44d	CG160131-03	363	364	Glycerol kinase (EC 2.7.1.30) (ATP:glycerol 3-phosphotransferase) (Glycerokinase) (GK)
45a	CG166282-01	365	366	Serine/threonine-protein kinase Chk1 (EC 2.7.1.-)
46a	CG170739-01	367	368	Pendrin (Sodium- independent chloride/iodide transporter)
47a	CG171632-01	369	370	Gamma-aminobutyric- acid receptor rho-1

				subunit precursor (GABA(A) receptor)
47b	CG171632-01	371	372	Gamma-aminobutyric- acid receptor rho-1 subunit precursor (GABA(A) receptor)
48a	CG173066-01	373	374	Aquaporin 7 (Aquaporin- 7 like) (Aquaporin adipose) (AQPap)
49a	CG173085-01	375	376	Similar to thyroid hormone receptor
49b	311531811	377	378	Similar to thyroid hormone receptor
50a	CG173095-01	379	380	Ubiquitin-protein ligase E3 Mdm2 (EC 6.3.2.-) (p53-binding protein Mdm2) (Oncoprotein Mdm2) (Double minute 2 protein) (Hdm2)
50b	CG173095-02	381	382	Ubiquitin-protein ligase E3 Mdm2 (EC 6.3.2.-) (p53-binding protein Mdm2) (Oncoprotein Mdm2) (Double minute 2 protein) (Hdm2)
51a	CG173173-01	383	384	Gamma-aminobutyric- acid receptor alpha-5 subunit precursor (GABA(A) receptor)
52a	CG51213-01	385	386	Sequence 3 from Patent WO0123561
52b	CG51213-07	387	388	Sequence 3 from Patent WO0123561
52c	CG51213-02	389	390	Sequence 3 from Patent WO0123561
52d	CG51213-03	391	392	Sequence 3 from Patent WO0123561
52e	CG51213-04	393	394	Sequence 3 from Patent WO0123561
52f	CG51213-05	395	396	Sequence 3 from Patent WO0123561
52g	CG51213-06	397	398	Sequence 3 from Patent WO0123561
53a	CG56155-01	399	400	Plasma kallikrein precursor (EC 3.4.21.34) (Plasma prekallikrein) (Kininogenin) (Fletcher factor)
53b	CG56155-02	401	402	Plasma kallikrein precursor (EC 3.4.21.34) (Plasma prekallikrein) (Kininogenin) (Fletcher factor)
53c	CG56155-03	403	404	Plasma kallikrein precursor (EC 3.4.21.34) (Plasma prekallikrein) (Kininogenin) (Fletcher factor)
54a	CG57191-01	405	406	Retinal short-chain

				dehydrogenase/reductase RETSDR1
54b	CG57191-03	407	408	Retinal short-chain dehydrogenase/reductase RETSDR1
54c	CG57191-02	409	410	Retinal short-chain dehydrogenase/reductase RETSDR1
55a	CG59595-01	411	412	Ribonuclease 6 precursor
55b	169728691	413	414	Ribonuclease 6 precursor
55c	169728707	415	416	Ribonuclease 6 precursor
55d	169728746	417	418	Ribonuclease 6 precursor
55e	CG59595-02	419	420	Ribonuclease 6 precursor
55f	CG59595-03	421	422	Ribonuclease 6 precursor
55g	CG59595-04	423	424	Ribonuclease 6 precursor
55h	CG59595-05	425	426	Ribonuclease 6 precursor
56a	CG92142-01	427	428	Glycerol-3-phosphate acyltransferase, mitochondrial precursor
56b	CG92142-02	429	430	Glycerol-3-phosphate acyltransferase, mitochondrial precursor
57a	CG95765-01	431	432	Hypothetical protein
57b	CG95765-02	433	434	Hypothetical protein
58a	CG97178-01	435	436	Tryptophan 2,3- dioxygenase (EC 1.13.11.11) (Tryptophan pyrrolase) (Tryptophanase) (Tryptophan oxygenase) (Tryptamin 2,3- dioxygenase) (TRPO)
58b	275481043	437	438	Tryptophan 2,3- dioxygenase (EC 1.13.11.11) (Tryptophan pyrrolase) (Tryptophanase) (Tryptophan oxygenase) (Tryptamin 2,3- dioxygenase) (TRPO)
58c	275481043	439	440	Diamine acetyltransferase (EC 2.3.1.57) (Spermidine/spermine N(1)- acetyltransferase) (SSAT) (Putrescine acetyltransferase)
59a	CG98102-01	441	442	Diamine acetyltransferase (EC 2.3.1.57) (Spermidine/spermine N(1)- acetyltransferase) (SSAT) (Putrescine acetyltransferase)
59b	CG98102-03	443	444	Diamine acetyltransferase (EC 2.3.1.57)

				(Spermidine/spermine N(1)- acetyltransferase) (SSAT) (Putrescine acetyltransferase)
59c	CG98102-02	445	446	Diamine acetyltransferase (EC 2.3.1.57) (Spermidine/spermine N(1)- acetyltransferase) (SSAT) (Putrescine acetyltransferase)
59d	CG98102-04	447	448	Diamine acetyltransferase (EC 2.3.1.57) (Spermidine/spermine N(1)- acetyltransferase) (SSAT) (Putrescine acetyltransferase)
59e	CG98102-05	449	450	Diamine acetyltransferase (EC 2.3.1.57) (Spermidine/spermine N(1)- acetyltransferase) (SSAT) (Putrescine acetyltransferase)
59f	CG98102-06	451	452	Diamine acetyltransferase (EC 2.3.1.57) (Spermidine/spermine N(1)- acetyltransferase) (SSAT) (Putrescine acetyltransferase)

Table A indicates the homology of NOVX polypeptides to known protein families. Thus, the nucleic acids and polypeptides, antibodies and related compounds according to the invention corresponding to a NOVX as identified in column 1 of Table A will be useful  
5 in therapeutic and diagnostic applications implicated in, for example, pathologies and disorders associated with the known protein families identified in column 5 of Table A.

Pathologies, diseases, disorders and condition and the like that are associated with NOVX sequences include, but are not limited to: *e.g.*, cardiomyopathy, atherosclerosis, hypertension, congenital heart defects, aortic stenosis, atrial septal defect (ASD),  
10 atrioventricular (A-V) canal defect, ductus arteriosus, pulmonary stenosis, subaortic stenosis, ventricular septal defect (VSD), valve diseases, tuberous sclerosis, scleroderma, obesity, metabolic disturbances associated with obesity, transplantation, adrenoleukodystrophy, congenital adrenal hyperplasia, prostate cancer, diabetes, metabolic disorders, neoplasm; adenocarcinoma, lymphoma, uterus cancer, fertility, hemophilia,  
15 hypercoagulation, idiopathic thrombocytopenic purpura, immunodeficiencies, graft versus host disease, AIDS, bronchial asthma, Crohn's disease; multiple sclerosis, treatment of Albright Hereditary Osteodystrophy, infectious disease, anorexia, cancer-associated

cachexia, cancer, neurodegenerative disorders, Alzheimer's Disease, Parkinson's Disorder, immune disorders, hematopoietic disorders, and the various dyslipidemias,] the metabolic syndrome X and wasting disorders associated with chronic diseases and various cancers, as well as conditions such as transplantation and fertility.

5       NOVX nucleic acids and their encoded polypeptides are useful in a variety of applications and contexts. The various NOVX nucleic acids and polypeptides according to the invention are useful as novel members of the protein families according to the presence of domains and sequence relatedness to previously described proteins. Additionally, NOVX nucleic acids and polypeptides can also be used to identify proteins that are  
10 members of the family to which the NOVX polypeptides belong.

Consistent with other known members of the family of proteins, identified in column 5 of Table A, the NOVX polypeptides of the present invention show homology to, and contain domains that are characteristic of, other members of such protein families. Details of the sequence relatedness and domain analysis for each NOVX are presented in  
15 Example A.

The NOVX nucleic acids and polypeptides can also be used to screen for molecules, which inhibit or enhance NOVX activity or function. Specifically, the nucleic acids and polypeptides according to the invention may be used as targets for the identification of small molecules that modulate or inhibit diseases associated with the protein families listed  
20 in Table A.

The NOVX nucleic acids and polypeptides are also useful for detecting specific cell types. Details of the expression analysis for each NOVX are presented in Example C. Accordingly, the NOVX nucleic acids, polypeptides, antibodies and related compounds according to the invention will have diagnostic and therapeutic applications in the detection  
25 of a variety of diseases with differential expression in normal vs. diseased tissues, *e.g.* detection of a variety of cancers.

Additional utilities for NOVX nucleic acids and polypeptides according to the invention are disclosed herein.

### 30   **NOVX clones**

NOVX nucleic acids and their encoded polypeptides are useful in a variety of applications and contexts. The various NOVX nucleic acids and polypeptides according to the invention are useful as novel members of the protein families according to the presence

of domains and sequence relatedness to previously described proteins. Additionally, NOVX nucleic acids and polypeptides can also be used to identify proteins that are members of the family to which the NOVX polypeptides belong.

The NOVX genes and their corresponding encoded proteins are useful for preventing, treating or ameliorating medical conditions, *e.g.*, by protein or gene therapy. Pathological conditions can be diagnosed by determining the amount of the new protein in a sample or by determining the presence of mutations in the new genes. Specific uses are described for each of the NOVX genes, based on the tissues in which they are most highly expressed. Uses include developing products for the diagnosis or treatment of a variety of diseases and disorders.

The NOVX nucleic acids and proteins of the invention are useful in potential diagnostic and therapeutic applications and as a research tool. These include serving as a specific or selective nucleic acid or protein diagnostic and/or prognostic marker, wherein the presence or amount of the nucleic acid or the protein are to be assessed, as well as potential therapeutic applications such as the following: (i) a protein therapeutic, (ii) a small molecule drug target, (iii) an antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), (iv) a nucleic acid useful in gene therapy (gene delivery/gene ablation), and (v) a composition promoting tissue regeneration *in vitro* and *in vivo* (vi) a biological defense weapon.

In one specific embodiment, the invention includes an isolated polypeptide comprising an amino acid sequence selected from the group consisting of: (a) a mature form of the amino acid sequence selected from the group consisting of SEQ ID NO: 2n, wherein n is an integer between 1 and 226; (b) a variant of a mature form of the amino acid sequence selected from the group consisting of SEQ ID NO: 2n, wherein n is an integer between 1 and 226, wherein any amino acid in the mature form is changed to a different amino acid, provided that no more than 15% of the amino acid residues in the sequence of the mature form are so changed; (c) an amino acid sequence selected from the group consisting of SEQ ID NO: 2n, wherein n is an integer between 1 and 226; (d) a variant of the amino acid sequence selected from the group consisting of SEQ ID NO: 2n, wherein n is an integer between 1 and 226 wherein any amino acid specified in the chosen sequence is changed to a different amino acid, provided that no more than 15% of the amino acid residues in the sequence are so changed; and (e) a fragment of any of (a) through (d).

In another specific embodiment, the invention includes an isolated nucleic acid molecule comprising a nucleic acid sequence encoding a polypeptide comprising an amino acid sequence selected from the group consisting of: (a) a mature form of the amino acid sequence given SEQ ID NO: 2n, wherein n is an integer between 1 and 226; (b) a variant of a mature form of the amino acid sequence selected from the group consisting of SEQ ID NO: 2n, wherein n is an integer between 1 and 226 wherein any amino acid in the mature form of the chosen sequence is changed to a different amino acid, provided that no more than 15% of the amino acid residues in the sequence of the mature form are so changed; (c) the amino acid sequence selected from the group consisting of SEQ ID NO: 2n, wherein n is an integer between 1 and 226; (d) a variant of the amino acid sequence selected from the group consisting of SEQ ID NO: 2n, wherein n is an integer between 1 and 226, in which any amino acid specified in the chosen sequence is changed to a different amino acid, provided that no more than 15% of the amino acid residues in the sequence are so changed; (e) a nucleic acid fragment encoding at least a portion of a polypeptide comprising the amino acid sequence selected from the group consisting of SEQ ID NO: 2n, wherein n is an integer between 1 and 226 or any variant of said polypeptide wherein any amino acid of the chosen sequence is changed to a different amino acid, provided that no more than 10% of the amino acid residues in the sequence are so changed; and (f) the complement of any of said nucleic acid molecules.

In yet another specific embodiment, the invention includes an isolated nucleic acid molecule, wherein said nucleic acid molecule comprises a nucleotide sequence selected from the group consisting of: (a) the nucleotide sequence selected from the group consisting of SEQ ID NO: 2n-1, wherein n is an integer between 1 and 226; (b) a nucleotide sequence wherein one or more nucleotides in the nucleotide sequence selected from the group consisting of SEQ ID NO: 2n-1, wherein n is an integer between 1 and 226 is changed from that selected from the group consisting of the chosen sequence to a different nucleotide provided that no more than 15% of the nucleotides are so changed; (c) a nucleic acid fragment of the sequence selected from the group consisting of SEQ ID NO: 2n-1, wherein n is an integer between 1 and 226; and (d) a nucleic acid fragment wherein one or more nucleotides in the nucleotide sequence selected from the group consisting of SEQ ID NO: 2n-1, wherein n is an integer between 1 and 226 is changed from that selected from the group consisting of the chosen sequence to a different nucleotide provided that no more than 15% of the nucleotides are so changed.

### NOVX Nucleic Acids and Polypeptides

One aspect of the invention pertains to isolated nucleic acid molecules that encode NOVX polypeptides or biologically active portions thereof. Also included in the invention are nucleic acid fragments sufficient for use as hybridization probes to identify

5 NOVX-encoding nucleic acids (*e.g.*, NOVX mRNAs) and fragments for use as PCR primers for the amplification and/or mutation of NOVX nucleic acid molecules. As used herein, the term “nucleic acid molecule” is intended to include DNA molecules (*e.g.*, cDNA or genomic DNA), RNA molecules (*e.g.*, mRNA), analogs of the DNA or RNA generated using nucleotide analogs, and derivatives, fragments and homologs thereof. The

10 nucleic acid molecule may be single-stranded or double-stranded, but preferably is comprised double-stranded DNA.

A NOVX nucleic acid can encode a mature NOVX polypeptide. As used herein, a “mature” form of a polypeptide or protein disclosed in the present invention is the product of a naturally occurring polypeptide or precursor form or proprotein. The naturally

15 occurring polypeptide, precursor or proprotein includes, by way of nonlimiting example, the full-length gene product encoded by the corresponding gene. Alternatively, it may be defined as the polypeptide, precursor or proprotein encoded by an ORF described herein. The product “mature” form arises, by way of nonlimiting example, as a result of one or more naturally occurring processing steps that may take place within the cell (*e.g.*, host

20 cell) in which the gene product arises. Examples of such processing steps leading to a “mature” form of a polypeptide or protein include the cleavage of the N-terminal methionine residue encoded by the initiation codon of an ORF, or the proteolytic cleavage of a signal peptide or leader sequence. Thus a mature form arising from a precursor polypeptide or protein that has residues 1 to N, where residue 1 is the N-terminal

25 methionine, would have residues 2 through N remaining after removal of the N-terminal methionine. Alternatively, a mature form arising from a precursor polypeptide or protein having residues 1 to N, in which an N-terminal signal sequence from residue 1 to residue M is cleaved, would have the residues from residue M+1 to residue N remaining. Further as used herein, a “mature” form of a polypeptide or protein may arise from a step of

30 post-translational modification other than a proteolytic cleavage event. Such additional processes include, by way of non-limiting example, glycosylation, myristylation or phosphorylation. In general, a mature polypeptide or protein may result from the operation of only one of these processes, or a combination of any of them.

The term "probe", as utilized herein, refers to nucleic acid sequences of variable length, preferably between at least about 10 nucleotides (nt), about 100 nt, or as many as approximately, *e.g.*, 6,000 nt, depending upon the specific use. Probes are used in the detection of identical, similar, or complementary nucleic acid sequences. Longer length probes are generally obtained from a natural or recombinant source, are highly specific, and much slower to hybridize than shorter-length oligomer probes. Probes may be single-stranded or double-stranded and designed to have specificity in PCR, membrane-based hybridization technologies, or ELISA-like technologies.

The term "isolated" nucleic acid molecule, as used herein, is a nucleic acid that is separated from other nucleic acid molecules which are present in the natural source of the nucleic acid. Preferably, an "isolated" nucleic acid is free of sequences which naturally flank the nucleic acid (*i.e.*, sequences located at the 5'- and 3'-termini of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid is derived. For example, in various embodiments, the isolated NOVX nucleic acid molecules can contain less than about 5 kb, 4 kb, 3 kb, 2 kb, 1 kb, 0.5 kb or 0.1 kb of nucleotide sequences which naturally flank the nucleic acid molecule in genomic DNA of the cell/tissue from which the nucleic acid is derived (*e.g.*, brain, heart, liver, spleen, *etc.*). Moreover, an "isolated" nucleic acid molecule, such as a cDNA molecule, can be substantially free of other cellular material, or culture medium, or of chemical precursors or other chemicals.

A nucleic acid molecule of the invention, *e.g.*, a nucleic acid molecule having the nucleotide sequence of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 226, or a complement of this nucleotide sequence, can be isolated using standard molecular biology techniques and the sequence information provided herein. Using all or a portion of the nucleic acid sequence of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 226, as a hybridization probe, NOVX molecules can be isolated using standard hybridization and cloning techniques (*e.g.*, as described in Sambrook, *et al.*, (eds.), MOLECULAR CLONING: A LABORATORY MANUAL 2<sup>nd</sup> Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989; and Ausubel, *et al.*, (eds.), CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, New York, NY, 1993.)

A nucleic acid of the invention can be amplified using cDNA, mRNA or alternatively, genomic DNA, as a template with appropriate oligonucleotide primers according to standard PCR amplification techniques. The nucleic acid so amplified can be cloned into an appropriate vector and characterized by DNA sequence analysis.

Furthermore, oligonucleotides corresponding to NOVX nucleotide sequences can be prepared by standard synthetic techniques, *e.g.*, using an automated DNA synthesizer.

As used herein, the term “oligonucleotide” refers to a series of linked nucleotide residues. A short oligonucleotide sequence may be based on, or designed from, a genomic or cDNA sequence and is used to amplify, confirm, or reveal the presence of an identical, similar or complementary DNA or RNA in a particular cell or tissue. Oligonucleotides comprise a nucleic acid sequence having about 10 nt, 50 nt, or 100 nt in length, preferably about 15 nt to 30 nt in length. In one embodiment of the invention, an oligonucleotide comprising a nucleic acid molecule less than 100 nt in length would further comprise at least 6 contiguous nucleotides of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 226, or a complement thereof. Oligonucleotides may be chemically synthesized and may also be used as probes.

In another embodiment, an isolated nucleic acid molecule of the invention comprises a nucleic acid molecule that is a complement of the nucleotide sequence shown in SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 226, or a portion of this nucleotide sequence (*e.g.*, a fragment that can be used as a probe or primer or a fragment encoding a biologically-active portion of a NOVX polypeptide). A nucleic acid molecule that is complementary to the nucleotide sequence of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 226, is one that is sufficiently complementary to the nucleotide sequence of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 226, that it can hydrogen bond with few or no mismatches to the nucleotide sequence shown in SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 226, thereby forming a stable duplex.

As used herein, the term “complementary” refers to Watson-Crick or Hoogsteen base pairing between nucleotides units of a nucleic acid molecule, and the term “binding” means the physical or chemical interaction between two polypeptides or compounds or associated polypeptides or compounds or combinations thereof. Binding includes ionic, non-ionic, van der Waals, hydrophobic interactions, and the like. A physical interaction can be either direct or indirect. Indirect interactions may be through or due to the effects of another polypeptide or compound. Direct binding refers to interactions that do not take place through, or due to, the effect of another polypeptide or compound, but instead are without other substantial chemical intermediates.

A “fragment” provided herein is defined as a sequence of at least 6 (contiguous) nucleic acids or at least 4 (contiguous) amino acids, a length sufficient to allow for specific

hybridization in the case of nucleic acids or for specific recognition of an epitope in the case of amino acids, and is at most some portion less than a full length sequence.

Fragments may be derived from any contiguous portion of a nucleic acid or amino acid sequence of choice.

- 5           A full-length NOVX clone is identified as containing an ATG translation start codon and an in-frame stop codon. Any disclosed NOVX nucleotide sequence lacking an ATG start codon therefore encodes a truncated C-terminal fragment of the respective NOVX polypeptide, and requires that the corresponding full-length cDNA extend in the 5' direction of the disclosed sequence. Any disclosed NOVX nucleotide sequence lacking an
- 10 in-frame stop codon similarly encodes a truncated N-terminal fragment of the respective NOVX polypeptide, and requires that the corresponding full-length cDNA extend in the 3' direction of the disclosed sequence.

- A "derivative" is a nucleic acid sequence or amino acid sequence formed from the native compounds either directly, by modification or partial substitution. An "analog" is a
- 15 nucleic acid sequence or amino acid sequence that has a structure similar to, but not identical to, the native compound, *e.g.* they differs from it in respect to certain components or side chains. Analogs may be synthetic or derived from a different evolutionary origin and may have a similar or opposite metabolic activity compared to wild type. A
- "homolog" is a nucleic acid sequence or amino acid sequence of a particular gene that is
- 20 derived from different species.

- Derivatives and analogs may be full length or other than full length. Derivatives or analogs of the nucleic acids or proteins of the invention include, but are not limited to, molecules comprising regions that are substantially homologous to the nucleic acids or proteins of the invention, in various embodiments, by at least about 70%, 80%, or 95%
- 25 identity (with a preferred identity of 80-95%) over a nucleic acid or amino acid sequence of identical size or when compared to an aligned sequence in which the alignment is done by a computer homology program known in the art, or whose encoding nucleic acid is capable of hybridizing to the complement of a sequence encoding the proteins under stringent, moderately stringent, or low stringent conditions. *See e.g.* Ausubel, *et al.*, CURRENT
- 30 PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, New York, NY, 1993, and below.

A "homologous nucleic acid sequence" or "homologous amino acid sequence," or variations thereof, refer to sequences characterized by a homology at the nucleotide level or

amino acid level as discussed above. Homologous nucleotide sequences include those sequences coding for isoforms of NOVX polypeptides. Isoforms can be expressed in different tissues of the same organism as a result of, for example, alternative splicing of RNA. Alternatively, isoforms can be encoded by different genes. In the invention,

5 homologous nucleotide sequences include nucleotide sequences encoding for a NOVX polypeptide of species other than humans, including, but not limited to: vertebrates, and thus can include, *e.g.*, frog, mouse, rat, rabbit, dog, cat, cow, horse, and other organisms. Homologous nucleotide sequences also include, but are not limited to, naturally occurring allelic variations and mutations of the nucleotide sequences set forth herein. A homologous

10 nucleotide sequence does not, however, include the exact nucleotide sequence encoding human NOVX protein. Homologous nucleic acid sequences include those nucleic acid sequences that encode conservative amino acid substitutions (see below) in SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 226, as well as a polypeptide possessing NOVX biological activity. Various biological activities of the NOVX proteins are

15 described below.

A NOVX polypeptide is encoded by the open reading frame ("ORF") of a NOVX nucleic acid. An ORF corresponds to a nucleotide sequence that could potentially be translated into a polypeptide. A stretch of nucleic acids comprising an ORF is uninterrupted by a stop codon. An ORF that represents the coding sequence for a full

20 protein begins with an ATG "start" codon and terminates with one of the three "stop" codons, namely, TAA, TAG, or TGA. For the purposes of this invention, an ORF may be any part of a coding sequence, with or without a start codon, a stop codon, or both. For an ORF to be considered as a good candidate for coding for a *bona fide* cellular protein, a minimum size requirement is often set, *e.g.*, a stretch of DNA that would encode a protein

25 of 50 amino acids or more.

The nucleotide sequences determined from the cloning of the human NOVX genes allows for the generation of probes and primers designed for use in identifying and/or cloning NOVX homologues in other cell types, *e.g.* from other tissues, as well as NOVX homologues from other vertebrates. The probe/primer typically comprises substantially

30 purified oligonucleotide. The oligonucleotide typically comprises a region of nucleotide sequence that hybridizes under stringent conditions to at least about 12, 25, 50, 100, 150, 200, 250, 300, 350 or 400 consecutive sense strand nucleotide sequence of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 226; or an anti-sense strand nucleotide

sequence of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 226; or of a naturally occurring mutant of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 226.

Probes based on the human NOVX nucleotide sequences can be used to detect transcripts or genomic sequences encoding the same or homologous proteins. In various  
 5 embodiments, the probe has a detectable label attached, *e.g.* the label can be a radioisotope, a fluorescent compound, an enzyme, or an enzyme co-factor. Such probes can be used as a part of a diagnostic test kit for identifying cells or tissues which mis-express a NOVX protein, such as by measuring a level of a NOVX-encoding nucleic acid in a sample of cells from a subject *e.g.*, detecting NOVX mRNA levels or determining whether a genomic  
 10 NOVX gene has been mutated or deleted.

"A polypeptide having a biologically-active portion of a NOVX polypeptide" refers to polypeptides exhibiting activity similar, but not necessarily identical to, an activity of a polypeptide of the invention, including mature forms, as measured in a particular biological assay, with or without dose dependency. A nucleic acid fragment encoding a  
 15 "biologically-active portion of NOVX" can be prepared by isolating a portion of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 226, that encodes a polypeptide having a NOVX biological activity (the biological activities of the NOVX proteins are described below), expressing the encoded portion of NOVX protein (*e.g.*, by recombinant expression *in vitro*) and assessing the activity of the encoded portion of NOVX.

20

#### **NOVX Nucleic Acid and Polypeptide Variants**

The invention further encompasses nucleic acid molecules that differ from the nucleotide sequences of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 226, due to degeneracy of the genetic code and thus encode the same NOVX proteins as that  
 25 encoded by the nucleotide sequences of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 226. In another embodiment, an isolated nucleic acid molecule of the invention has a nucleotide sequence encoding a protein having an amino acid sequence of SEQ ID NO:2*n*, wherein *n* is an integer between 1 and 226.

In addition to the human NOVX nucleotide sequences of SEQ ID NO:2*n*-1, wherein  
 30 *n* is an integer between 1 and 226, it will be appreciated by those skilled in the art that DNA sequence polymorphisms that lead to changes in the amino acid sequences of the NOVX polypeptides may exist within a population (*e.g.*, the human population). Such genetic polymorphism in the NOVX genes may exist among individuals within a

population due to natural allelic variation. As used herein, the terms "gene" and "recombinant gene" refer to nucleic acid molecules comprising an open reading frame (ORF) encoding a NOVX protein, preferably a vertebrate NOVX protein. Such natural allelic variations can typically result in 1-5% variance in the nucleotide sequence of the NOVX genes. Any and all such nucleotide variations and resulting amino acid polymorphisms in the NOVX polypeptides, which are the result of natural allelic variation and that do not alter the functional activity of the NOVX polypeptides, are intended to be within the scope of the invention.

Moreover, nucleic acid molecules encoding NOVX proteins from other species, and thus that have a nucleotide sequence that differs from a human SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 226, are intended to be within the scope of the invention. Nucleic acid molecules corresponding to natural allelic variants and homologues of the NOVX cDNAs of the invention can be isolated based on their homology to the human NOVX nucleic acids disclosed herein using the human cDNAs, or a portion thereof, as a hybridization probe according to standard hybridization techniques under stringent hybridization conditions.

Accordingly, in another embodiment, an isolated nucleic acid molecule of the invention is at least 6 nucleotides in length and hybridizes under stringent conditions to the nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 226. In another embodiment, the nucleic acid is at least 10, 25, 50, 100, 250, 500, 750, 1000, 1500, or 2000 or more nucleotides in length. In yet another embodiment, an isolated nucleic acid molecule of the invention hybridizes to the coding region. As used herein, the term "hybridizes under stringent conditions" is intended to describe conditions for hybridization and washing under which nucleotide sequences at least about 65% homologous to each other typically remain hybridized to each other.

Homologs (*i.e.*, nucleic acids encoding NOVX proteins derived from species other than human) or other related sequences (*e.g.*, paralogs) can be obtained by low, moderate or high stringency hybridization with all or a portion of the particular human sequence as a probe using methods well known in the art for nucleic acid hybridization and cloning.

As used herein, the phrase "stringent hybridization conditions" refers to conditions under which a probe, primer or oligonucleotide will hybridize to its target sequence, but to no other sequences. Stringent conditions are sequence-dependent and will be different in different circumstances. Longer sequences hybridize specifically at higher temperatures

than shorter sequences. Generally, stringent conditions are selected to be about 5 °C lower than the thermal melting point ( $T_m$ ) for the specific sequence at a defined ionic strength and pH. The  $T_m$  is the temperature (under defined ionic strength, pH and nucleic acid concentration) at which 50% of the probes complementary to the target sequence hybridize to the target sequence at equilibrium. Since the target sequences are generally present at excess, at  $T_m$ , 50% of the probes are occupied at equilibrium. Typically, stringent conditions will be those in which the salt concentration is less than about 1.0 M sodium ion, typically about 0.01 to 1.0 M sodium ion (or other salts) at pH 7.0 to 8.3 and the temperature is at least about 30 °C for short probes, primers or oligonucleotides (*e.g.*, 10 nt to 50 nt) and at least about 60 °C for longer probes, primers and oligonucleotides. Stringent conditions may also be achieved with the addition of destabilizing agents, such as formamide.

Stringent conditions are known to those skilled in the art and can be found in Ausubel, *et al.*, (eds.), CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6. Preferably, the conditions are such that sequences at least about 65%, 70%, 75%, 85%, 90%, 95%, 98%, or 99% homologous to each other typically remain hybridized to each other. A non-limiting example of stringent hybridization conditions are hybridization in a high salt buffer comprising 6X SSC, 50 mM Tris-HCl (pH 7.5), 1 mM EDTA, 0.02% PVP, 0.02% Ficoll, 0.02% BSA, and 500 mg/ml denatured salmon sperm DNA at 65°C, followed by one or more washes in 0.2X SSC, 0.01% BSA at 50°C. An isolated nucleic acid molecule of the invention that hybridizes under stringent conditions to a sequence of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 226, corresponds to a naturally-occurring nucleic acid molecule. As used herein, a "naturally-occurring" nucleic acid molecule refers to an RNA or DNA molecule having a nucleotide sequence that occurs in nature (*e.g.*, encodes a natural protein).

In a second embodiment, a nucleic acid sequence that is hybridizable to the nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 226, or fragments, analogs or derivatives thereof, under conditions of moderate stringency is provided. A non-limiting example of moderate stringency hybridization conditions are hybridization in 6X SSC, 5X Reinhardt's solution, 0.5% SDS and 100 mg/ml denatured salmon sperm DNA at 55 °C, followed by one or more washes in 1X SSC, 0.1% SDS at 37 °C. Other conditions of moderate stringency that may be used are well-known within the art. *See, e.g.*, Ausubel, *et al.* (eds.), 1993, CURRENT PROTOCOLS

IN MOLECULAR BIOLOGY, John Wiley & Sons, NY, and Krieger, 1990; GENE TRANSFER AND EXPRESSION, A LABORATORY MANUAL, Stockton Press, NY.

In a third embodiment, a nucleic acid that is hybridizable to the nucleic acid molecule comprising the nucleotide sequences of SEQ ID NO:2*n*-1, wherein *n* is an integer  
 5 between 1 and 226, or fragments, analogs or derivatives thereof, under conditions of low stringency, is provided. A non-limiting example of low stringency hybridization conditions are hybridization in 35% formamide, 5X SSC, 50 mM Tris-HCl (pH 7.5), 5 mM EDTA, 0.02% PVP, 0.02% Ficoll, 0.2% BSA, 100 mg/ml denatured salmon sperm DNA, 10% (wt/vol) dextran sulfate at 40°C, followed by one or more washes in 2X SSC, 25 mM  
 10 Tris-HCl (pH 7.4), 5 mM EDTA, and 0.1% SDS at 50°C. Other conditions of low stringency that may be used are well known in the art (*e.g.*, as employed for cross-species hybridizations). *See, e.g.*, Ausubel, *et al.* (eds.), 1993, CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, NY, and Kriegler, 1990, GENE TRANSFER AND EXPRESSION, A LABORATORY MANUAL, Stockton Press, NY; Shilo and Weinberg, 1981.  
 15 *Proc Natl Acad Sci USA* 78: 6789-6792.

### Conservative Mutations

In addition to naturally-occurring allelic variants of NOVX sequences that may exist in the population, the skilled artisan will further appreciate that changes can be  
 20 introduced by mutation into the nucleotide sequences of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 226, thereby leading to changes in the amino acid sequences of the encoded NOVX protein, without altering the functional ability of that NOVX protein. For example, nucleotide substitutions leading to amino acid substitutions at "non-essential" amino acid residues can be made in the sequence of SEQ ID NO:2*n*, wherein *n* is an integer  
 25 between 1 and 226. A "non-essential" amino acid residue is a residue that can be altered from the wild-type sequences of the NOVX proteins without altering their biological activity, whereas an "essential" amino acid residue is required for such biological activity. For example, amino acid residues that are conserved among the NOVX proteins of the invention are predicted to be particularly non-amenable to alteration. Amino acids for  
 30 which conservative substitutions can be made are well-known within the art.

Another aspect of the invention pertains to nucleic acid molecules encoding NOVX proteins that contain changes in amino acid residues that are not essential for activity. Such NOVX proteins differ in amino acid sequence from SEQ ID NO:2*n*-1, wherein *n* is an

integer between 1 and 226, yet retain biological activity. In one embodiment, the isolated nucleic acid molecule comprises a nucleotide sequence encoding a protein, wherein the protein comprises an amino acid sequence at least about 40% homologous to the amino acid sequences of SEQ ID NO:2*n*, wherein *n* is an integer between 1 and 226. Preferably, the protein encoded by the nucleic acid molecule is at least about 60% homologous to SEQ ID NO:2*n*, wherein *n* is an integer between 1 and 226; more preferably at least about 70% homologous to SEQ ID NO:2*n*, wherein *n* is an integer between 1 and 226; still more preferably at least about 80% homologous to SEQ ID NO:2*n*, wherein *n* is an integer between 1 and 226; even more preferably at least about 90% homologous to SEQ ID NO:2*n*, wherein *n* is an integer between 1 and 226; and most preferably at least about 95% homologous to SEQ ID NO:2*n*, wherein *n* is an integer between 1 and 226.

An isolated nucleic acid molecule encoding a NOVX protein homologous to the protein of SEQ ID NO:2*n*, wherein *n* is an integer between 1 and 226, can be created by introducing one or more nucleotide substitutions, additions or deletions into the nucleotide sequence of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 226, such that one or more amino acid substitutions, additions or deletions are introduced into the encoded protein.

Mutations can be introduced any one of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 226, by standard techniques, such as site-directed mutagenesis and PCR-mediated mutagenesis. Preferably, conservative amino acid substitutions are made at one or more predicted, non-essential amino acid residues. A "conservative amino acid substitution" is one in which the amino acid residue is replaced with an amino acid residue having a similar side chain. Families of amino acid residues having similar side chains have been defined within the art. These families include amino acids with basic side chains (*e.g.*, lysine, arginine, histidine), acidic side chains (*e.g.*, aspartic acid, glutamic acid), uncharged polar side chains (*e.g.*, glycine, asparagine, glutamine, serine, threonine, tyrosine, cysteine), nonpolar side chains (*e.g.*, alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan), beta-branched side chains (*e.g.*, threonine, valine, isoleucine) and aromatic side chains (*e.g.*, tyrosine, phenylalanine, tryptophan, histidine). Thus, a predicted non-essential amino acid residue in the NOVX protein is replaced with another amino acid residue from the same side chain family. Alternatively, in another embodiment, mutations can be introduced randomly along all or part of a NOVX coding sequence, such as by saturation mutagenesis, and the resultant mutants can be screened for

NOVX biological activity to identify mutants that retain activity. Following mutagenesis of a nucleic acid of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 226, the encoded protein can be expressed by any recombinant technology known in the art and the activity of the protein can be determined.

- 5           The relatedness of amino acid families may also be determined based on side chain interactions. Substituted amino acids may be fully conserved “strong” residues or fully conserved “weak” residues. The “strong” group of conserved amino acid residues may be any one of the following groups: STA, NEQK, NHQK, NDEQ, QHRK, MILV, MILF, HY, FYW, wherein the single letter amino acid codes are grouped by those amino acids that
- 10          may be substituted for each other. Likewise, the “weak” group of conserved residues may be any one of the following: CSA, ATV, SAG, STNK, STPA, SGND, SNDEQK, NDEQHK, NEQHRK, HFY, wherein the letters within each group represent the single letter amino acid code.

- In one embodiment, a mutant NOVX protein can be assayed for (i) the ability to
- 15          form protein:protein interactions with other NOVX proteins, other cell-surface proteins, or biologically-active portions thereof, (ii) complex formation between a mutant NOVX protein and a NOVX ligand; or (iii) the ability of a mutant NOVX protein to bind to an intracellular target protein or biologically-active portion thereof; (e.g. avidin proteins).

- In yet another embodiment, a mutant NOVX protein can be assayed for the ability
- 20          to regulate a specific biological function (e.g., regulation of insulin release).

### Interfering RNA

- In one aspect of the invention, NOVX gene expression can be attenuated by RNA interference. One approach well-known in the art is short interfering RNA (siRNA)
- 25          mediated gene silencing where expression products of a NOVX gene are targeted by specific double stranded NOVX derived siRNA nucleotide sequences that are complementary to at least a 19-25 nt long segment of the NOVX gene transcript, including the 5' untranslated (UT) region, the ORF, or the 3' UT region. *See, e.g.*, PCT applications WO00/44895, WO99/32619, WO01/75164, WO01/92513, WO 01/29058, WO01/89304,
- 30          WO02/16620, and WO02/29858, each incorporated by reference herein in their entirety. Targeted genes can be a NOVX gene, or an upstream or downstream modulator of the NOVX gene. Nonlimiting examples of upstream or downstream modulators of a NOVX gene include, e.g., a transcription factor that binds the NOVX gene promoter, a kinase or

phosphatase that interacts with a NOVX polypeptide, and polypeptides involved in a NOVX regulatory pathway.

According to the methods of the present invention, NOVX gene expression is silenced using short interfering RNA. A NOVX polynucleotide according to the invention includes a siRNA polynucleotide. Such a NOVX siRNA can be obtained using a NOVX polynucleotide sequence, for example, by processing the NOVX ribopolynucleotide sequence in a cell-free system, such as but not limited to a *Drosophila* extract, or by transcription of recombinant double stranded NOVX RNA or by chemical synthesis of nucleotide sequences homologous to a NOVX sequence. *See, e.g.*, Tuschl, Zamore, Lehmann, Bartel and Sharp (1999), *Genes & Dev.* 13: 3191-3197, incorporated herein by reference in its entirety. When synthesized, a typical 0.2 micromolar-scale RNA synthesis provides about 1 milligram of siRNA, which is sufficient for 1000 transfection experiments using a 24-well tissue culture plate format.

The most efficient silencing is generally observed with siRNA duplexes composed of a 21-nt sense strand and a 21-nt antisense strand, paired in a manner to have a 2-nt 3' overhang. The sequence of the 2-nt 3' overhang makes an additional small contribution to the specificity of siRNA target recognition. The contribution to specificity is localized to the unpaired nucleotide adjacent to the first paired bases. In one embodiment, the nucleotides in the 3' overhang are ribonucleotides. In an alternative embodiment, the nucleotides in the 3' overhang are deoxyribonucleotides. Using 2'-deoxyribonucleotides in the 3' overhangs is as efficient as using ribonucleotides, but deoxyribonucleotides are often cheaper to synthesize and are most likely more nuclease resistant.

A contemplated recombinant expression vector of the invention comprises a NOVX DNA molecule cloned into an expression vector comprising operatively-linked regulatory sequences flanking the NOVX sequence in a manner that allows for expression (by transcription of the DNA molecule) of both strands. An RNA molecule that is antisense to NOVX mRNA is transcribed by a first promoter (*e.g.*, a promoter sequence 3' of the cloned DNA) and an RNA molecule that is the sense strand for the NOVX mRNA is transcribed by a second promoter (*e.g.*, a promoter sequence 5' of the cloned DNA). The sense and antisense strands may hybridize *in vivo* to generate siRNA constructs for silencing of the NOVX gene. Alternatively, two constructs can be utilized to create the sense and antisense strands of a siRNA construct. Finally, cloned DNA can encode a construct having secondary structure, wherein a single transcript has both the sense and complementary

antisense sequences from the target gene or genes. In an example of this embodiment, a hairpin RNAi product is homologous to all or a portion of the target gene. In another example, a hairpin RNAi product is a siRNA. The regulatory sequences flanking the NOVX sequence may be identical or may be different, such that their expression may be modulated independently, or in a temporal or spatial manner.

In a specific embodiment, siRNAs are transcribed intracellularly by cloning the NOVX gene templates into a vector containing, *e.g.*, a RNA pol III transcription unit from the smaller nuclear RNA (snRNA) U6 or the human RNase P RNA H1. One example of a vector system is the GeneSuppressor<sup>TM</sup> RNA Interference kit (commercially available from Imgenex). The U6 and H1 promoters are members of the type III class of Pol III promoters. The +1 nucleotide of the U6-like promoters is always guanosine, whereas the +1 for H1 promoters is adenosine. The termination signal for these promoters is defined by five consecutive thymidines. The transcript is typically cleaved after the second uridine. Cleavage at this position generates a 3' UU overhang in the expressed siRNA, which is similar to the 3' overhangs of synthetic siRNAs. Any sequence less than 400 nucleotides in length can be transcribed by these promoter, therefore they are ideally suited for the expression of around 21-nucleotide siRNAs in, *e.g.*, an approximately 50-nucleotide RNA stem-loop transcript.

A siRNA vector appears to have an advantage over synthetic siRNAs where long term knock-down of expression is desired. Cells transfected with a siRNA expression vector would experience steady, long-term mRNA inhibition. In contrast, cells transfected with exogenous synthetic siRNAs typically recover from mRNA suppression within seven days or ten rounds of cell division. The long-term gene silencing ability of siRNA expression vectors may provide for applications in gene therapy.

In general, siRNAs are chopped from longer dsRNA by an ATP-dependent ribonuclease called DICER. DICER is a member of the RNase III family of double-stranded RNA-specific endonucleases. The siRNAs assemble with cellular proteins into an endonuclease complex. *In vitro* studies in *Drosophila* suggest that the siRNAs/protein complex (siRNP) is then transferred to a second enzyme complex, called an RNA-induced silencing complex (RISC), which contains an endoribonuclease that is distinct from DICER. RISC uses the sequence encoded by the antisense siRNA strand to find and destroy mRNAs of complementary sequence. The siRNA thus acts as a guide, restricting the ribonuclease to cleave only mRNAs complementary to one of the two siRNA strands.

A NOVX mRNA region to be targeted by siRNA is generally selected from a desired NOVX sequence beginning 50 to 100 nt downstream of the start codon. Alternatively, 5' or 3' UTRs and regions nearby the start codon can be used but are generally avoided, as these may be richer in regulatory protein binding sites. UTR-binding proteins and/or translation initiation complexes may interfere with binding of the siRNP or RISC endonuclease complex. An initial BLAST homology search for the selected siRNA sequence is done against an available nucleotide sequence library to ensure that only one gene is targeted. Specificity of target recognition by siRNA duplexes indicate that a single point mutation located in the paired region of an siRNA duplex is sufficient to abolish target mRNA degradation. See, Elbashir *et al.* 2001 EMBO J. 20(23):6877-88. Hence, consideration should be taken to accommodate SNPs, polymorphisms, allelic variants or species-specific variations when targeting a desired gene.

In one embodiment, a complete NOVX siRNA experiment includes the proper negative control. A negative control siRNA generally has the same nucleotide composition as the NOVX siRNA but lack significant sequence homology to the genome. Typically, one would scramble the nucleotide sequence of the NOVX siRNA and do a homology search to make sure it lacks homology to any other gene.

Two independent NOVX siRNA duplexes can be used to knock-down a target NOVX gene. This helps to control for specificity of the silencing effect. In addition, expression of two independent genes can be simultaneously knocked down by using equal concentrations of different NOVX siRNA duplexes, *e.g.*, a NOVX siRNA and an siRNA for a regulator of a NOVX gene or polypeptide. Availability of siRNA-associating proteins is believed to be more limiting than target mRNA accessibility.

A targeted NOVX region is typically a sequence of two adenines (AA) and two thymidines (TT) divided by a spacer region of nineteen (N19) residues (*e.g.*, AA(N19)TT). A desirable spacer region has a G/C-content of approximately 30% to 70%, and more preferably of about 50%. If the sequence AA(N19)TT is not present in the target sequence, an alternative target region would be AA(N21). The sequence of the NOVX sense siRNA corresponds to (N19)TT or N21, respectively. In the latter case, conversion of the 3' end of the sense siRNA to TT can be performed if such a sequence does not naturally occur in the NOVX polynucleotide. The rationale for this sequence conversion is to generate a symmetric duplex with respect to the sequence composition of the sense and antisense 3' overhangs. Symmetric 3' overhangs may help to ensure that the siRNPs are formed with

approximately equal ratios of sense and antisense target RNA-cleaving siRNPs. *See, e.g.,* Elbashir, Lendeckel and Tuschl (2001). *Genes & Dev.* 15: 188-200, incorporated by reference herein in its entirety. The modification of the overhang of the sense sequence of the siRNA duplex is not expected to affect targeted mRNA recognition, as the antisense  
5 siRNA strand guides target recognition.

Alternatively, if the NOVX target mRNA does not contain a suitable AA(N21) sequence, one may search for the sequence NA(N21). Further, the sequence of the sense strand and antisense strand may still be synthesized as 5' (N19)TT, as it is believed that the sequence of the 3'-most nucleotide of the antisense siRNA does not contribute to  
10 specificity. Unlike antisense or ribozyme technology, the secondary structure of the target mRNA does not appear to have a strong effect on silencing. *See, Harborth, et al.* (2001) *J. Cell Science* 114: 4557-4565, incorporated by reference in its entirety.

Transfection of NOVX siRNA duplexes can be achieved using standard nucleic acid transfection methods, for example, OLIGOFECTAMINE Reagent (commercially  
15 available from Invitrogen). An assay for NOVX gene silencing is generally performed approximately 2 days after transfection. No NOVX gene silencing has been observed in the absence of transfection reagent, allowing for a comparative analysis of the wild-type and silenced NOVX phenotypes. In a specific embodiment, for one well of a 24-well plate, approximately 0.84  $\mu$ g of the siRNA duplex is generally sufficient. Cells are typically  
20 seeded the previous day, and are transfected at about 50% confluence. The choice of cell culture media and conditions are routine to those of skill in the art, and will vary with the choice of cell type. The efficiency of transfection may depend on the cell type, but also on the passage number and the confluency of the cells. The time and the manner of formation of siRNA-liposome complexes (*e.g.* inversion versus vortexing) are also critical. Low  
25 transfection efficiencies are the most frequent cause of unsuccessful NOVX silencing. The efficiency of transfection needs to be carefully examined for each new cell line to be used. Preferred cell are derived from a mammal, more preferably from a rodent such as a rat or mouse, and most preferably from a human. Where used for therapeutic treatment, the cells are preferentially autologous, although non-autologous cell sources are also contemplated  
30 as within the scope of the present invention.

For a control experiment, transfection of 0.84  $\mu$ g single-stranded sense NOVX siRNA will have no effect on NOVX silencing, and 0.84  $\mu$ g antisense siRNA has a weak silencing effect when compared to 0.84  $\mu$ g of duplex siRNAs. Control experiments again

allow for a comparative analysis of the wild-type and silenced NOVX phenotypes. To control for transfection efficiency, targeting of common proteins is typically performed, for example targeting of lamin A/C or transfection of a CMV-driven EGFP-expression plasmid (*e.g.* commercially available from Clontech). In the above example, a determination of the fraction of lamin A/C knockdown in cells is determined the next day by such techniques as immunofluorescence, Western blot, Northern blot or other similar assays for protein expression or gene expression. Lamin A/C monoclonal antibodies may be obtained from Santa Cruz Biotechnology.

Depending on the abundance and the half life (or turnover) of the targeted NOVX polynucleotide in a cell, a knock-down phenotype may become apparent after 1 to 3 days, or even later. In cases where no NOVX knock-down phenotype is observed, depletion of the NOVX polynucleotide may be observed by immunofluorescence or Western blotting. If the NOVX polynucleotide is still abundant after 3 days, cells need to be split and transferred to a fresh 24-well plate for re-transfection. If no knock-down of the targeted protein is observed, it may be desirable to analyze whether the target mRNA (NOVX or a NOVX upstream or downstream gene) was effectively destroyed by the transfected siRNA duplex. Two days after transfection, total RNA is prepared, reverse transcribed using a target-specific primer, and PCR-amplified with a primer pair covering at least one exon-exon junction in order to control for amplification of pre-mRNAs. RT/PCR of a non-targeted mRNA is also needed as control. Effective depletion of the mRNA yet undetectable reduction of target protein may indicate that a large reservoir of stable NOVX protein may exist in the cell. Multiple transfection in sufficiently long intervals may be necessary until the target protein is finally depleted to a point where a phenotype may become apparent. If multiple transfection steps are required, cells are split 2 to 3 days after transfection. The cells may be transfected immediately after splitting.

An inventive therapeutic method of the invention contemplates administering a NOVX siRNA construct as therapy to compensate for increased or aberrant NOVX expression or activity. The NOVX ribopolynucleotide is obtained and processed into siRNA fragments, or a NOVX siRNA is synthesized, as described above. The NOVX siRNA is administered to cells or tissues using known nucleic acid transfection techniques, as described above. A NOVX siRNA specific for a NOVX gene will decrease or knockdown NOVX transcription products, which will lead to reduced NOVX polypeptide production, resulting in reduced NOVX polypeptide activity in the cells or tissues.

The present invention also encompasses a method of treating a disease or condition associated with the presence of a NOVX protein in an individual comprising administering to the individual an RNAi construct that targets the mRNA of the protein (the mRNA that encodes the protein) for degradation. A specific RNAi construct includes a siRNA or a  
5 double stranded gene transcript that is processed into siRNAs. Upon treatment, the target protein is not produced or is not produced to the extent it would be in the absence of the treatment.

Where the NOVX gene function is not correlated with a known phenotype, a control sample of cells or tissues from healthy individuals provides a reference standard for  
10 determining NOVX expression levels. Expression levels are detected using the assays described, *e.g.*, RT-PCR, Northern blotting, Western blotting, ELISA, and the like. A subject sample of cells or tissues is taken from a mammal, preferably a human subject, suffering from a disease state. The NOVX ribopolynucleotide is used to produce siRNA constructs, that are specific for the NOVX gene product. These cells or tissues are treated  
15 by administering NOVX siRNA's to the cells or tissues by methods described for the transfection of nucleic acids into a cell or tissue, and a change in NOVX polypeptide or polynucleotide expression is observed in the subject sample relative to the control sample, using the assays described. This NOVX gene knockdown approach provides a rapid method for determination of a NOVX minus (NOVX<sup>-</sup>) phenotype in the treated subject  
20 sample. The NOVX<sup>-</sup> phenotype observed in the treated subject sample thus serves as a marker for monitoring the course of a disease state during treatment.

In specific embodiments, a NOVX siRNA is used in therapy. Methods for the generation and use of a NOVX siRNA are known to those skilled in the art. Example techniques are provided below.

25

#### **Production of RNAs**

Sense RNA (ssRNA) and antisense RNA (asRNA) of NOVX are produced using known methods such as transcription in RNA expression vectors. In the initial experiments, the sense and antisense RNA are about 500 bases in length each. The  
30 produced ssRNA and asRNA (0.5  $\mu$ M) in 10 mM Tris-HCl (pH 7.5) with 20 mM NaCl were heated to 95° C for 1 min then cooled and annealed at room temperature for 12 to 16 h. The RNAs are precipitated and resuspended in lysis buffer (below). To monitor annealing, RNAs are electrophoresed in a 2% agarose gel in TBE buffer and stained with

ethidium bromide. See, *e.g.*, Sambrook et al., Molecular Cloning. Cold Spring Harbor Laboratory Press, Plainview, N.Y. (1989).

### Lysate Preparation

5           Untreated rabbit reticulocyte lysate (Ambion) are assembled according to the manufacturer's directions. dsRNA is incubated in the lysate at 30° C for 10 min prior to the addition of mRNAs. Then NOVX mRNAs are added and the incubation continued for an additional 60 min. The molar ratio of double stranded RNA and mRNA is about 200:1. The NOVX mRNA is radiolabeled (using known techniques) and its stability is monitored  
10 by gel electrophoresis.

          In a parallel experiment made with the same conditions, the double stranded RNA is internally radiolabeled with a <sup>32</sup>P-ATP. Reactions are stopped by the addition of 2 X proteinase K buffer and deproteinized as described previously (Tuschl *et al.*, Genes Dev., 13:3191-3197 (1999)). Products are analyzed by electrophoresis in 15% or 18%  
15 polyacrylamide sequencing gels using appropriate RNA standards. By monitoring the gels for radioactivity, the natural production of 10 to 25 nt RNAs from the double stranded RNA can be determined.

          The band of double stranded RNA, about 21-23 bps, is eluded. The efficacy of these 21-23 mers for suppressing NOVX transcription is assayed in vitro using the same  
20 rabbit reticulocyte assay described above using 50 nanomolar of double stranded 21-23 mer for each assay. The sequence of these 21-23 mers is then determined using standard nucleic acid sequencing techniques.

### RNA Preparation

25           21 nt RNAs, based on the sequence determined above, are chemically synthesized using Expedite RNA phosphoramidites and thymidine phosphoramidite (Proligo, Germany). Synthetic oligonucleotides are deprotected and gel-purified (Elbashir, Lendeckel, & Tuschl, Genes & Dev. 15, 188-200 (2001)), followed by Sep-Pak C18 cartridge (Waters, Milford, Mass., USA) purification (Tuschl, et al., Biochemistry,  
30 32:11658-11668 (1993)).

          These RNAs (20 μM) single strands are incubated in annealing buffer (100 mM potassium acetate, 30 mM HEPES-KOH at pH 7.4, 2 mM magnesium acetate) for 1 min at 90° C followed by 1 h at 37° C.

### Cell Culture

A cell culture known in the art to regularly express NOVX is propagated using standard conditions. 24 hours before transfection, at approx. 80% confluency, the cells are trypsinized and diluted 1:5 with fresh medium without antibiotics ( $1-3 \times 10^5$  cells/ml) and transferred to 24-well plates (500  $\mu$ l/well). Transfection is performed using a commercially available lipofection kit and NOVX expression is monitored using standard techniques with positive and negative control. A positive control is cells that naturally express NOVX while a negative control is cells that do not express NOVX. Base-paired 21 and 22 nt siRNAs with overhanging 3' ends mediate efficient sequence-specific mRNA degradation in lysates and in cell culture. Different concentrations of siRNAs are used. An efficient concentration for suppression in vitro in mammalian culture is between 25 nM to 100 nM final concentration. This indicates that siRNAs are effective at concentrations that are several orders of magnitude below the concentrations applied in conventional antisense or ribozyme gene targeting experiments.

The above method provides a way both for the deduction of NOVX siRNA sequence and the use of such siRNA for in vitro suppression. In vivo suppression may be performed using the same siRNA using well known in vivo transfection or gene therapy transfection techniques.

### Antisense Nucleic Acids

Another aspect of the invention pertains to isolated antisense nucleic acid molecules that are hybridizable to or complementary to the nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 226, or fragments, analogs or derivatives thereof. An "antisense" nucleic acid comprises a nucleotide sequence that is complementary to a "sense" nucleic acid encoding a protein (*e.g.*, complementary to the coding strand of a double-stranded cDNA molecule or complementary to an mRNA sequence). In specific aspects, antisense nucleic acid molecules are provided that comprise a sequence complementary to at least about 10, 25, 50, 100, 250 or 500 nucleotides or an entire NOVX coding strand, or to only a portion thereof. Nucleic acid molecules encoding fragments, homologs, derivatives and analogs of a NOVX protein of SEQ ID NO:2*n*, wherein *n* is an integer between 1 and 226, or antisense nucleic acids complementary to a NOVX nucleic acid sequence of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 226, are additionally provided.

In one embodiment, an antisense nucleic acid molecule is antisense to a "coding region" of the coding strand of a nucleotide sequence encoding a NOVX protein. The term "coding region" refers to the region of the nucleotide sequence comprising codons which are translated into amino acid residues. In another embodiment, the antisense nucleic acid molecule is antisense to a "noncoding region" of the coding strand of a nucleotide sequence encoding the NOVX protein. The term "noncoding region" refers to 5' and 3' sequences which flank the coding region that are not translated into amino acids (*i.e.*, also referred to as 5' and 3' untranslated regions).

Given the coding strand sequences encoding the NOVX protein disclosed herein, antisense nucleic acids of the invention can be designed according to the rules of Watson and Crick or Hoogsteen base pairing. The antisense nucleic acid molecule can be complementary to the entire coding region of NOVX mRNA, but more preferably is an oligonucleotide that is antisense to only a portion of the coding or noncoding region of NOVX mRNA. For example, the antisense oligonucleotide can be complementary to the region surrounding the translation start site of NOVX mRNA. An antisense oligonucleotide can be, for example, about 5, 10, 15, 20, 25, 30, 35, 40, 45 or 50 nucleotides in length. An antisense nucleic acid of the invention can be constructed using chemical synthesis or enzymatic ligation reactions using procedures known in the art. For example, an antisense nucleic acid (*e.g.*, an antisense oligonucleotide) can be chemically synthesized using naturally-occurring nucleotides or variously modified nucleotides designed to increase the biological stability of the molecules or to increase the physical stability of the duplex formed between the antisense and sense nucleic acids (*e.g.*, phosphorothioate derivatives and acridine substituted nucleotides can be used).

Examples of modified nucleotides that can be used to generate the antisense nucleic acid include: 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xanthine, 4-acetylcytosine, 5-carboxymethylaminomethyl-2-thiouridine, 5-(carboxyhydroxymethyl) uracil, 5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 5-methoxyuracil, 3-methylcytosine, 5-methylcytosine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, 2-thiouracil, 4-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarboxymethyluracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid (v), wybutoxosine,

pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 5-methyluracil, uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid (v), 5-methyl-2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl) uracil, (acp3)w, and 2,6-diaminopurine. Alternatively, the antisense nucleic acid can be produced biologically using an expression vector into  
5 which a nucleic acid has been subcloned in an antisense orientation (*i.e.*, RNA transcribed from the inserted nucleic acid will be of an antisense orientation to a target nucleic acid of interest, described further in the following subsection).

The antisense nucleic acid molecules of the invention are typically administered to a subject or generated *in situ* such that they hybridize with or bind to cellular mRNA and/or  
10 genomic DNA encoding a NOVX protein to thereby inhibit expression of the protein (*e.g.*, by inhibiting transcription and/or translation). The hybridization can be by conventional nucleotide complementarity to form a stable duplex, or, for example, in the case of an antisense nucleic acid molecule that binds to DNA duplexes, through specific interactions in the major groove of the double helix. An example of a route of administration of  
15 antisense nucleic acid molecules of the invention includes direct injection at a tissue site. Alternatively, antisense nucleic acid molecules can be modified to target selected cells and then administered systemically. For example, for systemic administration, antisense molecules can be modified such that they specifically bind to receptors or antigens expressed on a selected cell surface (*e.g.*, by linking the antisense nucleic acid molecules to  
20 peptides or antibodies that bind to cell surface receptors or antigens). The antisense nucleic acid molecules can also be delivered to cells using the vectors described herein. To achieve sufficient nucleic acid molecules, vector constructs in which the antisense nucleic acid molecule is placed under the control of a strong pol II or pol III promoter are preferred.

In yet another embodiment, the antisense nucleic acid molecule of the invention is  
25 an  $\alpha$ -anomeric nucleic acid molecule. An  $\alpha$ -anomeric nucleic acid molecule forms specific double-stranded hybrids with complementary RNA in which, contrary to the usual  $\beta$ -units, the strands run parallel to each other. *See, e.g.*, Gaultier, *et al.*, 1987. *Nucl. Acids Res.* **15**: 6625-6641. The antisense nucleic acid molecule can also comprise a 2'-o-methylribonucleotide (*See, e.g.*, Inoue, *et al.* 1987. *Nucl. Acids Res.* **15**: 6131-6148) or  
30 a chimeric RNA-DNA analogue (*See, e.g.*, Inoue, *et al.*, 1987. *FEBS Lett.* **215**: 327-330).

#### Ribozymes and PNA Moieties

Nucleic acid modifications include, by way of non-limiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding  
5 nucleic acids in therapeutic applications in a subject.

In one embodiment, an antisense nucleic acid of the invention is a ribozyme. Ribozymes are catalytic RNA molecules with ribonuclease activity that are capable of cleaving a single-stranded nucleic acid, such as an mRNA, to which they have a complementary region. Thus, ribozymes (*e.g.*, hammerhead ribozymes as described in  
10 Haselhoff and Gerlach 1988. *Nature* 334: 585-591) can be used to catalytically cleave NOVX mRNA transcripts to thereby inhibit translation of NOVX mRNA. A ribozyme having specificity for a NOVX-encoding nucleic acid can be designed based upon the nucleotide sequence of a NOVX cDNA disclosed herein (*i.e.*, SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 226). For example, a derivative of a *Tetrahymena* L-19 IVS  
15 RNA can be constructed in which the nucleotide sequence of the active site is complementary to the nucleotide sequence to be cleaved in a NOVX-encoding mRNA. *See, e.g.*, U.S. Patent 4,987,071 to Cech, *et al.* and U.S. Patent 5,116,742 to Cech, *et al.* NOVX mRNA can also be used to select a catalytic RNA having a specific ribonuclease activity from a pool of RNA molecules. *See, e.g.*, Bartel *et al.*, (1993) *Science*  
20 261:1411-1418.

Alternatively, NOVX gene expression can be inhibited by targeting nucleotide sequences complementary to the regulatory region of the NOVX nucleic acid (*e.g.*, the NOVX promoter and/or enhancers) to form triple helical structures that prevent transcription of the NOVX gene in target cells. *See, e.g.*, Helene, 1991. *Anticancer Drug*  
25 *Des.* 6: 569-84; Helene, *et al.* 1992. *Ann. N.Y. Acad. Sci.* 660: 27-36; Maher, 1992. *Bioassays* 14: 807-15.

In various embodiments, the NOVX nucleic acids can be modified at the base moiety, sugar moiety or phosphate backbone to improve, *e.g.*, the stability, hybridization, or solubility of the molecule. For example, the deoxyribose phosphate backbone of the  
30 nucleic acids can be modified to generate peptide nucleic acids. *See, e.g.*, Hyrup, *et al.*, 1996. *Bioorg Med. Chem* 4: 5-23. As used herein, the terms "peptide nucleic acids" or "PNAs" refer to nucleic acid mimics (*e.g.*, DNA mimics) in which the deoxyribose phosphate backbone is replaced by a pseudopeptide backbone and only the four natural

nucleotide bases are retained. The neutral backbone of PNAs has been shown to allow for specific hybridization to DNA and RNA under conditions of low ionic strength. The synthesis of PNA oligomer can be performed using standard solid phase peptide synthesis protocols as described in Hyrup, *et al.*, 1996. *supra*; Perry-O'Keefe, *et al.*, 1996. *Proc. Natl. Acad. Sci. USA* 93: 14670-14675.

PNAs of NOVX can be used in therapeutic and diagnostic applications. For example, PNAs can be used as antisense or antigene agents for sequence-specific modulation of gene expression by, *e.g.*, inducing transcription or translation arrest or inhibiting replication. PNAs of NOVX can also be used, for example, in the analysis of single base pair mutations in a gene (*e.g.*, PNA directed PCR clamping; as artificial restriction enzymes when used in combination with other enzymes, *e.g.*, S<sub>1</sub> nucleases (*See*, Hyrup, *et al.*, 1996. *supra*); or as probes or primers for DNA sequence and hybridization (*See*, Hyrup, *et al.*, 1996, *supra*; Perry-O'Keefe, *et al.*, 1996. *supra*).

In another embodiment, PNAs of NOVX can be modified, *e.g.*, to enhance their stability or cellular uptake, by attaching lipophilic or other helper groups to PNA, by the formation of PNA-DNA chimeras, or by the use of liposomes or other techniques of drug delivery known in the art. For example, PNA-DNA chimeras of NOVX can be generated that may combine the advantageous properties of PNA and DNA. Such chimeras allow DNA recognition enzymes (*e.g.*, RNase H and DNA polymerases) to interact with the DNA portion while the PNA portion would provide high binding affinity and specificity. PNA-DNA chimeras can be linked using linkers of appropriate lengths selected in terms of base stacking, number of bonds between the nucleotide bases, and orientation (*see*, Hyrup, *et al.*, 1996. *supra*). The synthesis of PNA-DNA chimeras can be performed as described in Hyrup, *et al.*, 1996. *supra* and Finn, *et al.*, 1996. *Nucl Acids Res* 24: 3357-3363. For example, a DNA chain can be synthesized on a solid support using standard phosphoramidite coupling chemistry, and modified nucleoside analogs, *e.g.*, 5'-(4-methoxytrityl)amino-5'-deoxy-thymidine phosphoramidite, can be used between the PNA and the 5' end of DNA. *See, e.g.*, Mag, *et al.*, 1989. *Nucl Acid Res* 17: 5973-5988. PNA monomers are then coupled in a stepwise manner to produce a chimeric molecule with a 5' PNA segment and a 3' DNA segment. *See, e.g.*, Finn, *et al.*, 1996. *supra*. Alternatively, chimeric molecules can be synthesized with a 5' DNA segment and a 3' PNA segment. *See, e.g.*, Petersen, *et al.*, 1975. *Bioorg. Med. Chem. Lett.* 5: 1119-11124.

In other embodiments, the oligonucleotide may include other appended groups such as peptides (*e.g.*, for targeting host cell receptors *in vivo*), or agents facilitating transport across the cell membrane (*see, e.g.*, Letsinger, *et al.*, 1989. *Proc. Natl. Acad. Sci. U.S.A.* 86: 6553-6556; Lemaitre, *et al.*, 1987. *Proc. Natl. Acad. Sci.* 84: 648-652; PCT Publication No. WO88/09810) or the blood-brain barrier (*see, e.g.*, PCT Publication No. WO 89/10134). In addition, oligonucleotides can be modified with hybridization triggered cleavage agents (*see, e.g.*, Krol, *et al.*, 1988. *BioTechniques* 6:958-976) or intercalating agents (*see, e.g.*, Zon, 1988. *Pharm. Res.* 5: 539-549). To this end, the oligonucleotide may be conjugated to another molecule, *e.g.*, a peptide, a hybridization triggered cross-linking agent, a transport agent, a hybridization-triggered cleavage agent, and the like.

### NOVX Polypeptides

A polypeptide according to the invention includes a polypeptide including the amino acid sequence of NOVX polypeptides whose sequences are provided in any one of SEQ ID NO:2*n*, wherein *n* is an integer between 1 and 226. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residues shown in any one of SEQ ID NO:2*n*, wherein *n* is an integer between 1 and 226, while still encoding a protein that maintains its NOVX activities and physiological functions, or a functional fragment thereof.

In general, a NOVX variant that preserves NOVX-like function includes any variant in which residues at a particular position in the sequence have been substituted by other amino acids, and further include the possibility of inserting an additional residue or residues between two residues of the parent protein as well as the possibility of deleting one or more residues from the parent sequence. Any amino acid substitution, insertion, or deletion is encompassed by the invention. In favorable circumstances, the substitution is a conservative substitution as defined above.

One aspect of the invention pertains to isolated NOVX proteins, and biologically-active portions thereof, or derivatives, fragments, analogs or homologs thereof. Also provided are polypeptide fragments suitable for use as immunogens to raise anti-NOVX antibodies. In one embodiment, native NOVX proteins can be isolated from cells or tissue sources by an appropriate purification scheme using standard protein purification techniques. In another embodiment, NOVX proteins are produced by

recombinant DNA techniques. Alternative to recombinant expression, a NOVX protein or polypeptide can be synthesized chemically using standard peptide synthesis techniques.

An "isolated" or "purified" polypeptide or protein or biologically-active portion thereof is substantially free of cellular material or other contaminating proteins from the cell or tissue source from which the NOVX protein is derived, or substantially free from chemical precursors or other chemicals when chemically synthesized. The language "substantially free of cellular material" includes preparations of NOVX proteins in which the protein is separated from cellular components of the cells from which it is isolated or recombinantly-produced. In one embodiment, the language "substantially free of cellular material" includes preparations of NOVX proteins having less than about 30% (by dry weight) of non-NOVX proteins (also referred to herein as a "contaminating protein"), more preferably less than about 20% of non-NOVX proteins, still more preferably less than about 10% of non-NOVX proteins, and most preferably less than about 5% of non-NOVX proteins. When the NOVX protein or biologically-active portion thereof is recombinantly-produced, it is also preferably substantially free of culture medium, *i.e.*, culture medium represents less than about 20%, more preferably less than about 10%, and most preferably less than about 5% of the volume of the NOVX protein preparation.

The language "substantially free of chemical precursors or other chemicals" includes preparations of NOVX proteins in which the protein is separated from chemical precursors or other chemicals that are involved in the synthesis of the protein. In one embodiment, the language "substantially free of chemical precursors or other chemicals" includes preparations of NOVX proteins having less than about 30% (by dry weight) of chemical precursors or non-NOVX chemicals, more preferably less than about 20% chemical precursors or non-NOVX chemicals, still more preferably less than about 10% chemical precursors or non-NOVX chemicals, and most preferably less than about 5% chemical precursors or non-NOVX chemicals.

Biologically-active portions of NOVX proteins include peptides comprising amino acid sequences sufficiently homologous to or derived from the amino acid sequences of the NOVX proteins (*e.g.*, the amino acid sequence of SEQ ID NO:2*n*, wherein *n* is an integer between 1 and 226) that include fewer amino acids than the full-length NOVX proteins, and exhibit at least one activity of a NOVX protein. Typically, biologically-active portions comprise a domain or motif with at least one activity of the NOVX protein. A

biologically-active portion of a NOVX protein can be a polypeptide which is, for example, 10, 25, 50, 100 or more amino acid residues in length.

Moreover, other biologically-active portions, in which other regions of the protein are deleted, can be prepared by recombinant techniques and evaluated for one or more of the functional activities of a native NOVX protein.

In an embodiment, the NOVX protein has an amino acid sequence of SEQ ID NO:2*n*, wherein *n* is an integer between 1 and 226. In other embodiments, the NOVX protein is substantially homologous to SEQ ID NO:2*n*, wherein *n* is an integer between 1 and 226, and retains the functional activity of the protein of SEQ ID NO:2*n*, wherein *n* is an integer between 1 and 226, yet differs in amino acid sequence due to natural allelic variation or mutagenesis, as described in detail, below. Accordingly, in another embodiment, the NOVX protein is a protein that comprises an amino acid sequence at least about 45% homologous to the amino acid sequence of SEQ ID NO:2*n*, wherein *n* is an integer between 1 and 226, and retains the functional activity of the NOVX proteins of SEQ ID NO:2*n*, wherein *n* is an integer between 1 and 226.

### Determining Homology Between Two or More Sequences

To determine the percent homology of two amino acid sequences or of two nucleic acids, the sequences are aligned for optimal comparison purposes (*e.g.*, gaps can be introduced in the sequence of a first amino acid or nucleic acid sequence for optimal alignment with a second amino or nucleic acid sequence). The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in the first sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in the second sequence, then the molecules are homologous at that position (*i.e.*, as used herein amino acid or nucleic acid "homology" is equivalent to amino acid or nucleic acid "identity").

The nucleic acid sequence homology may be determined as the degree of identity between two sequences. The homology may be determined using computer programs known in the art, such as GAP software provided in the GCG program package. *See*, Needleman and Wunsch, 1970. *J Mol Biol* 48: 443-453. Using GCG GAP software with the following settings for nucleic acid sequence comparison: GAP creation penalty of 5.0 and GAP extension penalty of 0.3, the coding region of the analogous nucleic acid sequences referred to above exhibits a degree of identity preferably of at least 70%, 75%,

80%, 85%, 90%, 95%, 98%, or 99%, with the CDS (encoding) part of the DNA sequence of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 226.

The term "sequence identity" refers to the degree to which two polynucleotide or polypeptide sequences are identical on a residue-by-residue basis over a particular region of comparison. The term "percentage of sequence identity" is calculated by comparing two optimally aligned sequences over that region of comparison, determining the number of positions at which the identical nucleic acid base (*e.g.*, A, T, C, G, U, or I, in the case of nucleic acids) occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the region of comparison (*i.e.*, the window size), and multiplying the result by 100 to yield the percentage of sequence identity. The term "substantial identity" as used herein denotes a characteristic of a polynucleotide sequence, wherein the polynucleotide comprises a sequence that has at least 80 percent sequence identity, preferably at least 85 percent identity and often 90 to 95 percent sequence identity, more usually at least 99 percent sequence identity as compared to a reference sequence over a comparison region.

#### Chimeric and Fusion Proteins

The invention also provides NOVX chimeric or fusion proteins. As used herein, a NOVX "chimeric protein" or "fusion protein" comprises a NOVX polypeptide operatively-linked to a non-NOVX polypeptide. An "NOVX polypeptide" refers to a polypeptide having an amino acid sequence corresponding to a NOVX protein of SEQ ID NO:2*n*, wherein *n* is an integer between 1 and 226, whereas a "non-NOVX polypeptide" refers to a polypeptide having an amino acid sequence corresponding to a protein that is not substantially homologous to the NOVX protein, *e.g.*, a protein that is different from the NOVX protein and that is derived from the same or a different organism. Within a NOVX fusion protein the NOVX polypeptide can correspond to all or a portion of a NOVX protein. In one embodiment, a NOVX fusion protein comprises at least one biologically-active portion of a NOVX protein. In another embodiment, a NOVX fusion protein comprises at least two biologically-active portions of a NOVX protein. In yet another embodiment, a NOVX fusion protein comprises at least three biologically-active portions of a NOVX protein. Within the fusion protein, the term "operatively-linked" is intended to indicate that the NOVX polypeptide and the non-NOVX polypeptide are fused

in-frame with one another. The non-NOVX polypeptide can be fused to the N-terminus or C-terminus of the NOVX polypeptide.

In one embodiment, the fusion protein is a GST-NOVX fusion protein in which the NOVX sequences are fused to the C-terminus of the GST (glutathione S-transferase) sequences. Such fusion proteins can facilitate the purification of recombinant NOVX polypeptides.

In another embodiment, the fusion protein is a NOVX protein containing a heterologous signal sequence at its N-terminus. In certain host cells (*e.g.*, mammalian host cells), expression and/or secretion of NOVX can be increased through use of a heterologous signal sequence.

In yet another embodiment, the fusion protein is a NOVX-immunoglobulin fusion protein in which the NOVX sequences are fused to sequences derived from a member of the immunoglobulin protein family. The NOVX-immunoglobulin fusion proteins of the invention can be incorporated into pharmaceutical compositions and administered to a subject to inhibit an interaction between a NOVX ligand and a NOVX protein on the surface of a cell, to thereby suppress NOVX-mediated signal transduction *in vivo*. The NOVX-immunoglobulin fusion proteins can be used to affect the bioavailability of a NOVX cognate ligand. Inhibition of the NOVX ligand/NOVX interaction may be useful therapeutically for both the treatment of proliferative and differentiative disorders, as well as modulating (*e.g.* promoting or inhibiting) cell survival. Moreover, the NOVX-immunoglobulin fusion proteins of the invention can be used as immunogens to produce anti-NOVX antibodies in a subject, to purify NOVX ligands, and in screening assays to identify molecules that inhibit the interaction of NOVX with a NOVX ligand.

A NOVX chimeric or fusion protein of the invention can be produced by standard recombinant DNA techniques. For example, DNA fragments coding for the different polypeptide sequences are ligated together in-frame in accordance with conventional techniques, *e.g.*, by employing blunt-ended or stagger-ended termini for ligation, restriction enzyme digestion to provide for appropriate termini, filling-in of cohesive ends as appropriate, alkaline phosphatase treatment to avoid undesirable joining, and enzymatic ligation. In another embodiment, the fusion gene can be synthesized by conventional techniques including automated DNA synthesizers. Alternatively, PCR amplification of gene fragments can be carried out using anchor primers that give rise to complementary overhangs between two consecutive gene fragments that can subsequently be annealed and

reamplified to generate a chimeric gene sequence (*see, e.g.,* Ausubel, *et al.* (eds.) CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, 1992). Moreover, many expression vectors are commercially available that already encode a fusion moiety (*e.g.,* a GST polypeptide). A NOVX-encoding nucleic acid can be cloned into such an expression  
5 vector such that the fusion moiety is linked in-frame to the NOVX protein.

### NOVX Agonists and Antagonists

The invention also pertains to variants of the NOVX proteins that function as either NOVX agonists (*i.e.,* mimetics) or as NOVX antagonists. Variants of the NOVX protein  
10 can be generated by mutagenesis (*e.g.,* discrete point mutation or truncation of the NOVX protein). An agonist of the NOVX protein can retain substantially the same, or a subset of, the biological activities of the naturally occurring form of the NOVX protein. An antagonist of the NOVX protein can inhibit one or more of the activities of the naturally occurring form of the NOVX protein by, for example, competitively binding to a  
15 downstream or upstream member of a cellular signaling cascade which includes the NOVX protein. Thus, specific biological effects can be elicited by treatment with a variant of limited function. In one embodiment, treatment of a subject with a variant having a subset of the biological activities of the naturally occurring form of the protein has fewer side effects in a subject relative to treatment with the naturally occurring form of the NOVX  
20 proteins.

Variants of the NOVX proteins that function as either NOVX agonists (*i.e.,* mimetics) or as NOVX antagonists can be identified by screening combinatorial libraries of mutants (*e.g.,* truncation mutants) of the NOVX proteins for NOVX protein agonist or antagonist activity. In one embodiment, a variegated library of NOVX variants is  
25 generated by combinatorial mutagenesis at the nucleic acid level and is encoded by a variegated gene library. A variegated library of NOVX variants can be produced by, for example, enzymatically ligating a mixture of synthetic oligonucleotides into gene sequences such that a degenerate set of potential NOVX sequences is expressible as individual polypeptides, or alternatively, as a set of larger fusion proteins (*e.g.,* for phage  
30 display) containing the set of NOVX sequences therein. There are a variety of methods which can be used to produce libraries of potential NOVX variants from a degenerate oligonucleotide sequence. Chemical synthesis of a degenerate gene sequence can be performed in an automatic DNA synthesizer, and the synthetic gene then ligated into an

appropriate expression vector. Use of a degenerate set of genes allows for the provision, in one mixture, of all of the sequences encoding the desired set of potential NOVX sequences. Methods for synthesizing degenerate oligonucleotides are well-known within the art. *See, e.g.,* Narang, 1983. *Tetrahedron* 39: 3; Itakura, *et al.*, 1984. *Annu. Rev. Biochem.* 53: 323; 5 Itakura, *et al.*, 1984. *Science* 198: 1056; Ike, *et al.*, 1983. *Nucl. Acids Res.* 11: 477.

### Polypeptide Libraries

In addition, libraries of fragments of the NOVX protein coding sequences can be used to generate a variegated population of NOVX fragments for screening and subsequent 10 selection of variants of a NOVX protein. In one embodiment, a library of coding sequence fragments can be generated by treating a double stranded PCR fragment of a NOVX coding sequence with a nuclease under conditions wherein nicking occurs only about once per molecule, denaturing the double stranded DNA, renaturing the DNA to form double-stranded DNA that can include sense/antisense pairs from different nicked products, 15 removing single stranded portions from reformed duplexes by treatment with S<sub>1</sub> nuclease, and ligating the resulting fragment library into an expression vector. By this method, expression libraries can be derived which encodes N-terminal and internal fragments of various sizes of the NOVX proteins.

Various techniques are known in the art for screening gene products of 20 combinatorial libraries made by point mutations or truncation, and for screening cDNA libraries for gene products having a selected property. Such techniques are adaptable for rapid screening of the gene libraries generated by the combinatorial mutagenesis of NOVX proteins. The most widely used techniques, which are amenable to high throughput analysis, for screening large gene libraries typically include cloning the gene library into 25 replicable expression vectors, transforming appropriate cells with the resulting library of vectors, and expressing the combinatorial genes under conditions in which detection of a desired activity facilitates isolation of the vector encoding the gene whose product was detected. Recursive ensemble mutagenesis (REM), a new technique that enhances the frequency of functional mutants in the libraries, can be used in combination with the 30 screening assays to identify NOVX variants. *See, e.g.,* Arkin and Yourvan, 1992. *Proc. Natl. Acad. Sci. USA* 89: 7811-7815; Delgrave, *et al.*, 1993. *Protein Engineering* 6:327-331.

### Anti-NOVX Antibodies

- Included in the invention are antibodies to NOVX proteins, or fragments of NOVX proteins. The term "antibody" as used herein refers to immunoglobulin molecules and immunologically active portions of immunoglobulin (Ig) molecules, *i.e.*, molecules that
- 5 contain an antigen binding site that specifically binds (immunoreacts with) an antigen. Such antibodies include, but are not limited to, polyclonal, monoclonal, chimeric, single chain, F<sub>ab</sub>, F<sub>ab</sub>' and F<sub>(ab)</sub>'<sub>2</sub> fragments, and an F<sub>ab</sub> expression library. In general, antibody molecules obtained from humans relates to any of the classes IgG, IgM, IgA, IgE and IgD, which differ from one another by the nature of the heavy chain present in the molecule.
- 10 Certain classes have subclasses as well, such as IgG<sub>1</sub>, IgG<sub>2</sub>, and others. Furthermore, in humans, the light chain may be a kappa chain or a lambda chain. Reference herein to antibodies includes a reference to all such classes, subclasses and types of human antibody species.

- An isolated protein of the invention intended to serve as an antigen, or a portion or
- 15 fragment thereof, can be used as an immunogen to generate antibodies that immunospecifically bind the antigen, using standard techniques for polyclonal and monoclonal antibody preparation. The full-length protein can be used or, alternatively, the invention provides antigenic peptide fragments of the antigen for use as immunogens. An antigenic peptide fragment comprises at least 6 amino acid residues of the amino acid
- 20 sequence of the full length protein, such as an amino acid sequence of SEQ ID NO:2*n*, wherein *n* is an integer between 1 and 226, and encompasses an epitope thereof such that an antibody raised against the peptide forms a specific immune complex with the full length protein or with any fragment that contains the epitope. Preferably, the antigenic peptide comprises at least 10 amino acid residues, or at least 15 amino acid residues, or at
- 25 least 20 amino acid residues, or at least 30 amino acid residues. Preferred epitopes encompassed by the antigenic peptide are regions of the protein that are located on its surface; commonly these are hydrophilic regions.

- In certain embodiments of the invention, at least one epitope encompassed by the antigenic peptide is a region of NOVX that is located on the surface of the protein, *e.g.*, a
- 30 hydrophilic region. A hydrophobicity analysis of the human NOVX protein sequence will indicate which regions of a NOVX polypeptide are particularly hydrophilic and, therefore, are likely to encode surface residues useful for targeting antibody production. As a means for targeting antibody production, hydropathy plots showing regions of hydrophilicity and

hydrophobicity may be generated by any method well known in the art, including, for example, the Kyte Doolittle or the Hopp Woods methods, either with or without Fourier transformation. See, *e.g.*, Hopp and Woods, 1981, *Proc. Nat. Acad. Sci. USA* 78: 3824-3828; Kyte and Doolittle 1982, *J. Mol. Biol.* 157: 105-142, each incorporated herein  
5 by reference in their entirety. Antibodies that are specific for one or more domains within an antigenic protein, or derivatives, fragments, analogs or homologs thereof, are also provided herein.

The term "epitope" includes any protein determinant capable of specific binding to an immunoglobulin or T-cell receptor. Epitopic determinants usually consist of chemically  
10 active surface groupings of molecules such as amino acids or sugar side chains and usually have specific three dimensional structural characteristics, as well as specific charge characteristics. A NOVX polypeptide or a fragment thereof comprises at least one antigenic epitope. An anti-NOVX antibody of the present invention is said to specifically bind to antigen NOVX when the equilibrium binding constant ( $K_D$ ) is  $\leq 1 \mu\text{M}$ , preferably  $\leq 100$   
15 nM, more preferably  $\leq 10 \text{ nM}$ , and most preferably  $\leq 100 \text{ pM}$  to about  $1 \text{ pM}$ , as measured by assays such as radioligand binding assays or similar assays known to those skilled in the art.

A protein of the invention, or a derivative, fragment, analog, homolog or ortholog thereof, may be utilized as an immunogen in the generation of antibodies that  
20 immunospecifically bind these protein components.

Various procedures known within the art may be used for the production of polyclonal or monoclonal antibodies directed against a protein of the invention, or against derivatives, fragments, analogs homologs or orthologs thereof (see, for example, *Antibodies: A Laboratory Manual*, Harlow E, and Lane D, 1988, Cold Spring Harbor  
25 Laboratory Press, Cold Spring Harbor, NY, incorporated herein by reference). Some of these antibodies are discussed below.

### **Polyclonal Antibodies**

For the production of polyclonal antibodies, various suitable host animals (*e.g.*,  
30 rabbit, goat, mouse or other mammal) may be immunized by one or more injections with the native protein, a synthetic variant thereof, or a derivative of the foregoing. An appropriate immunogenic preparation can contain, for example, the naturally occurring immunogenic protein, a chemically synthesized polypeptide representing the immunogenic

protein, or a recombinantly expressed immunogenic protein. Furthermore, the protein may be conjugated to a second protein known to be immunogenic in the mammal being immunized. Examples of such immunogenic proteins include but are not limited to keyhole limpet hemocyanin, serum albumin, bovine thyroglobulin, and soybean trypsin inhibitor. The preparation can further include an adjuvant. Various adjuvants used to increase the immunological response include, but are not limited to, Freund's (complete and incomplete), mineral gels (*e.g.*, aluminum hydroxide), surface active substances (*e.g.*, lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, dinitrophenol, *etc.*), adjuvants usable in humans such as Bacille Calmette-Guerin and *Corynebacterium parvum*, or similar immunostimulatory agents. Additional examples of adjuvants which can be employed include MPL-TDM adjuvant (monophosphoryl Lipid A, synthetic trehalose dicorynomycolate).

The polyclonal antibody molecules directed against the immunogenic protein can be isolated from the mammal (*e.g.*, from the blood) and further purified by well known techniques, such as affinity chromatography using protein A or protein G, which provide primarily the IgG fraction of immune serum. Subsequently, or alternatively, the specific antigen which is the target of the immunoglobulin sought, or an epitope thereof, may be immobilized on a column to purify the immune specific antibody by immunoaffinity chromatography. Purification of immunoglobulins is discussed, for example, by D. Wilkinson (The Scientist, published by The Scientist, Inc., Philadelphia PA, Vol. 14, No. 8 (April 17, 2000), pp. 25-28).

### Monoclonal Antibodies

The term "monoclonal antibody" (MAb) or "monoclonal antibody composition", as used herein, refers to a population of antibody molecules that contain only one molecular species of antibody molecule consisting of a unique light chain gene product and a unique heavy chain gene product. In particular, the complementarity determining regions (CDRs) of the monoclonal antibody are identical in all the molecules of the population. MAbs thus contain an antigen binding site capable of immunoreacting with a particular epitope of the antigen characterized by a unique binding affinity for it.

Monoclonal antibodies can be prepared using hybridoma methods, such as those described by Kohler and Milstein, *Nature*, 256:495 (1975). In a hybridoma method, a mouse, hamster, or other appropriate host animal, is typically immunized with an

immunizing agent to elicit lymphocytes that produce or are capable of producing antibodies that will specifically bind to the immunizing agent. Alternatively, the lymphocytes can be immunized in vitro.

The immunizing agent will typically include the protein antigen, a fragment thereof  
5 or a fusion protein thereof. Generally, either peripheral blood lymphocytes are used if cells of human origin are desired, or spleen cells or lymph node cells are used if non-human mammalian sources are desired. The lymphocytes are then fused with an immortalized cell line using a suitable fusing agent, such as polyethylene glycol, to form a hybridoma cell (Goding, Monoclonal Antibodies: Principles and Practice, Academic Press, (1986) pp.  
10 59-103). Immortalized cell lines are usually transformed mammalian cells, particularly myeloma cells of rodent, bovine and human origin. Usually, rat or mouse myeloma cell lines are employed. The hybridoma cells can be cultured in a suitable culture medium that preferably contains one or more substances that inhibit the growth or survival of the unfused, immortalized cells. For example, if the parental cells lack the enzyme  
15 hypoxanthine guanine phosphoribosyl transferase (HGPRT or HPRT), the culture medium for the hybridomas typically will include hypoxanthine, aminopterin, and thymidine ("HAT medium"), which substances prevent the growth of HGPRT-deficient cells.

Preferred immortalized cell lines are those that fuse efficiently, support stable high level expression of antibody by the selected antibody-producing cells, and are sensitive to a  
20 medium such as HAT medium. More preferred immortalized cell lines are murine myeloma lines, which can be obtained, for instance, from the Salk Institute Cell Distribution Center, San Diego, California and the American Type Culture Collection, Manassas, Virginia. Human myeloma and mouse-human heteromyeloma cell lines also have been described for the production of human monoclonal antibodies (Kozbor, J.  
25 Immunol., 133:3001 (1984); Brodeur et al., Monoclonal Antibody Production Techniques and Applications, Marcel Dekker, Inc., New York, (1987) pp. 51-63).

The culture medium in which the hybridoma cells are cultured can then be assayed for the presence of monoclonal antibodies directed against the antigen. Preferably, the binding specificity of monoclonal antibodies produced by the hybridoma cells is  
30 determined by immunoprecipitation or by an in vitro binding assay, such as radioimmunoassay (RIA) or enzyme-linked immunoabsorbent assay (ELISA). Such techniques and assays are known in the art. The binding affinity of the monoclonal antibody can, for example, be determined by the Scatchard analysis of Munson and Pollard,

Anal. Biochem., 107:220 (1980). It is an objective, especially important in therapeutic applications of monoclonal antibodies, to identify antibodies having a high degree of specificity and a high binding affinity for the target antigen.

After the desired hybridoma cells are identified, the clones can be subcloned by limiting dilution procedures and grown by standard methods (Goding, 1986). Suitable culture media for this purpose include, for example, Dulbecco's Modified Eagle's Medium and RPMI-1640 medium. Alternatively, the hybridoma cells can be grown in vivo as ascites in a mammal.

The monoclonal antibodies secreted by the subclones can be isolated or purified from the culture medium or ascites fluid by conventional immunoglobulin purification procedures such as, for example, protein A-Sepharose, hydroxylapatite chromatography, gel electrophoresis, dialysis, or affinity chromatography.

The monoclonal antibodies can also be made by recombinant DNA methods, such as those described in U.S. Patent No. 4,816,567. DNA encoding the monoclonal antibodies of the invention can be readily isolated and sequenced using conventional procedures (*e.g.*, by using oligonucleotide probes that are capable of binding specifically to genes encoding the heavy and light chains of murine antibodies). The hybridoma cells of the invention serve as a preferred source of such DNA. Once isolated, the DNA can be placed into expression vectors, which are then transfected into host cells such as simian COS cells, Chinese hamster ovary (CHO) cells, or myeloma cells that do not otherwise produce immunoglobulin protein, to obtain the synthesis of monoclonal antibodies in the recombinant host cells. The DNA also can be modified, for example, by substituting the coding sequence for human heavy and light chain constant domains in place of the homologous murine sequences (U.S. Patent No. 4,816,567; Morrison, Nature 368, 812-13 (1994)) or by covalently joining to the immunoglobulin coding sequence all or part of the coding sequence for a non-immunoglobulin polypeptide. Such a non-immunoglobulin polypeptide can be substituted for the constant domains of an antibody of the invention, or can be substituted for the variable domains of one antigen-combining site of an antibody of the invention to create a chimeric bivalent antibody.

30

#### **Humanized Antibodies**

The antibodies directed against the protein antigens of the invention can further comprise humanized antibodies or human antibodies. These antibodies are suitable for

administration to humans without engendering an immune response by the human against the administered immunoglobulin. Humanized forms of antibodies are chimeric immunoglobulins, immunoglobulin chains or fragments thereof (such as Fv, Fab, Fab', F(ab')<sub>2</sub> or other antigen-binding subsequences of antibodies) that are principally comprised of the sequence of a human immunoglobulin, and contain minimal sequence derived from a non-human immunoglobulin. Humanization can be performed following the method of Winter and co-workers (Jones et al., Nature, 321:522-525 (1986); Riechmann et al., Nature, 332:323-327 (1988); Verhoeyen et al., Science, 239:1534-1536 (1988)), by substituting rodent CDRs or CDR sequences for the corresponding sequences of a human antibody. (See also U.S. Patent No. 5,225,539.) In some instances, Fv framework residues of the human immunoglobulin are replaced by corresponding non-human residues. Humanized antibodies can also comprise residues which are found neither in the recipient antibody nor in the imported CDR or framework sequences. In general, the humanized antibody will comprise substantially all of at least one, and typically two, variable domains, in which all or substantially all of the CDR regions correspond to those of a non-human immunoglobulin and all or substantially all of the framework regions are those of a human immunoglobulin consensus sequence. The humanized antibody optimally also will comprise at least a portion of an immunoglobulin constant region (Fc), typically that of a human immunoglobulin (Jones et al., 1986; Riechmann et al., 1988; and Presta, Curr. Op. Struct. Biol., 2:593-596 (1992)).

### Human Antibodies

Fully human antibodies essentially relate to antibody molecules in which the entire sequence of both the light chain and the heavy chain, including the CDRs, arise from human genes. Such antibodies are termed "human antibodies", or "fully human antibodies" herein. Human monoclonal antibodies can be prepared by the trioma technique; the human B-cell hybridoma technique (see Kozbor, et al., 1983 Immunol Today 4: 72) and the EBV hybridoma technique to produce human monoclonal antibodies (see Cole, et al., 1985 In: MONOCLONAL ANTIBODIES AND CANCER THERAPY, Alan R. Liss, Inc., pp. 77-96). Human monoclonal antibodies may be utilized in the practice of the present invention and may be produced by using human hybridomas (see Cote, et al., 1983. Proc Natl Acad Sci USA 80: 2026-2030) or by transforming human B-cells with Epstein Barr Virus in vitro (see Cole, et

al., 1985 In: MONOCLONAL ANTIBODIES AND CANCER THERAPY, Alan R. Liss, Inc., pp. 77-96).

In addition, human antibodies can also be produced using additional techniques, including phage display libraries (Hoogenboom and Winter, J. Mol. Biol., 227:381 (1991); Marks et al., J. Mol. Biol., 222:581 (1991)). Similarly, human antibodies can be made by introducing human immunoglobulin loci into transgenic animals, *e.g.*, mice in which the endogenous immunoglobulin genes have been partially or completely inactivated. Upon challenge, human antibody production is observed, which closely resembles that seen in humans in all respects, including gene rearrangement, assembly, and antibody repertoire. This approach is described, for example, in U.S. Patent Nos. 5,545,807; 5,545,806; 5,569,825; 5,625,126; 5,633,425; 5,661,016, and in Marks et al. (Bio/Technology 10, 779-783 (1992)); Lonberg et al. (Nature 368 856-859 (1994)); Morrison (Nature 368, 812-13 (1994)); Fishwild et al. (Nature Biotechnology 14, 845-51 (1996)); Neuberger (Nature Biotechnology 14, 826 (1996)); and Lonberg and Huszar (Intern. Rev. Immunol. 13 65-93 (1995)).

Human antibodies may additionally be produced using transgenic nonhuman animals which are modified so as to produce fully human antibodies rather than the animal's endogenous antibodies in response to challenge by an antigen. (See PCT publication WO94/02602). The endogenous genes encoding the heavy and light immunoglobulin chains in the nonhuman host have been incapacitated, and active loci encoding human heavy and light chain immunoglobulins are inserted into the host's genome. The human genes are incorporated, for example, using yeast artificial chromosomes containing the requisite human DNA segments. An animal which provides all the desired modifications is then obtained as progeny by crossbreeding intermediate transgenic animals containing fewer than the full complement of the modifications. The preferred embodiment of such a nonhuman animal is a mouse, and is termed the Xenomouse<sup>TM</sup> as disclosed in PCT publications WO 96/33735 and WO 96/34096. This animal produces B cells which secrete fully human immunoglobulins. The antibodies can be obtained directly from the animal after immunization with an immunogen of interest, as, for example, a preparation of a polyclonal antibody, or alternatively from immortalized B cells derived from the animal, such as hybridomas producing monoclonal antibodies. Additionally, the genes encoding the immunoglobulins with human variable regions can be

recovered and expressed to obtain the antibodies directly, or can be further modified to obtain analogs of antibodies such as, for example, single chain Fv molecules.

An example of a method of producing a nonhuman host, exemplified as a mouse, lacking expression of an endogenous immunoglobulin heavy chain is disclosed in U.S. Patent No. 5,939,598. It can be obtained by a method including deleting the J segment genes from at least one endogenous heavy chain locus in an embryonic stem cell to prevent rearrangement of the locus and to prevent formation of a transcript of a rearranged immunoglobulin heavy chain locus, the deletion being effected by a targeting vector containing a gene encoding a selectable marker; and producing from the embryonic stem cell a transgenic mouse whose somatic and germ cells contain the gene encoding the selectable marker.

A method for producing an antibody of interest, such as a human antibody, is disclosed in U.S. Patent No. 5,916,771. It includes introducing an expression vector that contains a nucleotide sequence encoding a heavy chain into one mammalian host cell in culture, introducing an expression vector containing a nucleotide sequence encoding a light chain into another mammalian host cell, and fusing the two cells to form a hybrid cell. The hybrid cell expresses an antibody containing the heavy chain and the light chain.

In a further improvement on this procedure, a method for identifying a clinically relevant epitope on an immunogen, and a correlative method for selecting an antibody that binds immunospecifically to the relevant epitope with high affinity, are disclosed in PCT publication WO 99/53049.

#### **F<sub>ab</sub> Fragments and Single Chain Antibodies**

According to the invention, techniques can be adapted for the production of single-chain antibodies specific to an antigenic protein of the invention (see *e.g.*, U.S. Patent No. 4,946,778). In addition, methods can be adapted for the construction of F<sub>ab</sub> expression libraries (see *e.g.*, Huse, et al., 1989 Science 246: 1275-1281) to allow rapid and effective identification of monoclonal F<sub>ab</sub> fragments with the desired specificity for a protein or derivatives, fragments, analogs or homologs thereof. Antibody fragments that contain the idiotypes to a protein antigen may be produced by techniques known in the art including, but not limited to: (i) an F<sub>(ab)<sub>2</sub></sub> fragment produced by pepsin digestion of an antibody molecule; (ii) an F<sub>ab</sub> fragment generated by reducing the disulfide bridges of an

$F_{(ab)_2}$  fragment; (iii) an  $F_{ab}$  fragment generated by the treatment of the antibody molecule with papain and a reducing agent and (iv)  $F_v$  fragments.

### **Bispecific Antibodies**

5           Bispecific antibodies are monoclonal, preferably human or humanized, antibodies that have binding specificities for at least two different antigens. In the present case, one of the binding specificities is for an antigenic protein of the invention. The second binding target is any other antigen, and advantageously is a cell-surface protein or receptor or receptor subunit.

10           Methods for making bispecific antibodies are known in the art. Traditionally, the recombinant production of bispecific antibodies is based on the co-expression of two immunoglobulin heavy-chain/light-chain pairs, where the two heavy chains have different specificities (Milstein and Cuello, *Nature*, 305:537-539 (1983)). Because of the random assortment of immunoglobulin heavy and light chains, these hybridomas (quadromas)  
15           produce a potential mixture of ten different antibody molecules, of which only one has the correct bispecific structure. The purification of the correct molecule is usually accomplished by affinity chromatography steps. Similar procedures are disclosed in WO 93/08829, published 13 May 1993, and in Traunecker et al., *EMBO J.*, 10:3655-3659 (1991).

20           Antibody variable domains with the desired binding specificities (antibody-antigen combining sites) can be fused to immunoglobulin constant domain sequences. The fusion preferably is with an immunoglobulin heavy-chain constant domain, comprising at least part of the hinge, CH2, and CH3 regions. It is preferred to have the first heavy-chain constant region (CH1) containing the site necessary for light-chain binding present in at  
25           least one of the fusions. DNAs encoding the immunoglobulin heavy-chain fusions and, if desired, the immunoglobulin light chain, are inserted into separate expression vectors, and are co-transfected into a suitable host organism. For further details of generating bispecific antibodies see, for example, Suresh et al., *Methods in Enzymology*, 121:210 (1986).

          According to another approach described in WO 96/27011, the interface between a  
30           pair of antibody molecules can be engineered to maximize the percentage of heterodimers which are recovered from recombinant cell culture. The preferred interface comprises at least a part of the CH3 region of an antibody constant domain. In this method, one or more small amino acid side chains from the interface of the first antibody molecule are replaced

with larger side chains (*e.g.* tyrosine or tryptophan). Compensatory “cavities” of identical or similar size to the large side chain(s) are created on the interface of the second antibody molecule by replacing large amino acid side chains with smaller ones (*e.g.* alanine or threonine). This provides a mechanism for increasing the yield of the heterodimer over  
5 other unwanted end-products such as homodimers.

Bispecific antibodies can be prepared as full length antibodies or antibody fragments (*e.g.* F(ab')<sub>2</sub> bispecific antibodies). Techniques for generating bispecific antibodies from antibody fragments have been described in the literature. For example, bispecific antibodies can be prepared using chemical linkage. Brennan et al., Science  
10 229:81 (1985) describe a procedure wherein intact antibodies are proteolytically cleaved to generate F(ab')<sub>2</sub> fragments. These fragments are reduced in the presence of the dithiol complexing agent sodium arsenite to stabilize vicinal dithiols and prevent intermolecular disulfide formation. The Fab' fragments generated are then converted to thionitrobenzoate (TNB) derivatives. One of the Fab'-TNB derivatives is then reconverted to the Fab'-thiol  
15 by reduction with mercaptoethylamine and is mixed with an equimolar amount of the other Fab'-TNB derivative to form the bispecific antibody. The bispecific antibodies produced can be used as agents for the selective immobilization of enzymes.

Additionally, Fab' fragments can be directly recovered from *E. coli* and chemically coupled to form bispecific antibodies. Shalaby et al., J. Exp. Med. 175:217-225 (1992)  
20 describe the production of a fully humanized bispecific antibody F(ab')<sub>2</sub> molecule. Each Fab' fragment was separately secreted from *E. coli* and subjected to directed chemical coupling in vitro to form the bispecific antibody. The bispecific antibody thus formed was able to bind to cells overexpressing the ErbB2 receptor and normal human T cells, as well as trigger the lytic activity of human cytotoxic lymphocytes against human breast tumor  
25 targets.

Various techniques for making and isolating bispecific antibody fragments directly from recombinant cell culture have also been described. For example, bispecific antibodies have been produced using leucine zippers. Kostelny et al., J. Immunol. 148(5):1547-1553 (1992). The leucine zipper peptides from the Fos and Jun proteins were linked to the Fab'  
30 portions of two different antibodies by gene fusion. The antibody homodimers were reduced at the hinge region to form monomers and then re-oxidized to form the antibody heterodimers. This method can also be utilized for the production of antibody homodimers. The “diabody” technology described by Hollinger et al., Proc. Natl. Acad. Sci. USA

90:6444-6448 (1993) has provided an alternative mechanism for making bispecific antibody fragments. The fragments comprise a heavy-chain variable domain ( $V_H$ ) connected to a light-chain variable domain ( $V_L$ ) by a linker which is too short to allow pairing between the two domains on the same chain. Accordingly, the  $V_H$  and  $V_L$  domains of one fragment are forced to pair with the complementary  $V_L$  and  $V_H$  domains of another fragment, thereby forming two antigen-binding sites. Another strategy for making bispecific antibody fragments by the use of single-chain Fv (sFv) dimers has also been reported. See, Gruber et al., J. Immunol. 152:5368 (1994).

Antibodies with more than two valencies are contemplated. For example, trispecific antibodies can be prepared. Tutt et al., J. Immunol. 147:60 (1991).

Exemplary bispecific antibodies can bind to two different epitopes, at least one of which originates in the protein antigen of the invention. Alternatively, an anti-antigenic arm of an immunoglobulin molecule can be combined with an arm which binds to a triggering molecule on a leukocyte such as a T-cell receptor molecule (*e.g.* CD2, CD3, CD28, or B7), or Fc receptors for IgG ( $Fc\gamma R$ ), such as  $Fc\gamma RI$  (CD64),  $Fc\gamma RII$  (CD32) and  $Fc\gamma RIII$  (CD16) so as to focus cellular defense mechanisms to the cell expressing the particular antigen. Bispecific antibodies can also be used to direct cytotoxic agents to cells which express a particular antigen. These antibodies possess an antigen-binding arm and an arm which binds a cytotoxic agent or a radionuclide chelator, such as EOTUBE, DPTA, DOTA, or TETA. Another bispecific antibody of interest binds the protein antigen described herein and further binds tissue factor (TF).

### Heteroconjugate Antibodies

Heteroconjugate antibodies are also within the scope of the present invention. Heteroconjugate antibodies are composed of two covalently joined antibodies. Such antibodies have, for example, been proposed to target immune system cells to unwanted cells (U.S. Patent No. 4,676,980), and for treatment of HIV infection (WO 91/00360; WO 92/200373; EP 03089). It is contemplated that the antibodies can be prepared in vitro using known methods in synthetic protein chemistry, including those involving crosslinking agents. For example, immunotoxins can be constructed using a disulfide exchange reaction or by forming a thioether bond. Examples of suitable reagents for this purpose include iminothiolate and methyl-4-mercaptobutyrimidate and those disclosed, for example, in U.S. Patent No. 4,676,980.

### Effector Function Engineering

It can be desirable to modify the antibody of the invention with respect to effector function, so as to enhance, *e.g.*, the effectiveness of the antibody in treating cancer. For example, cysteine residue(s) can be introduced into the Fc region, thereby allowing

5 interchain disulfide bond formation in this region. The homodimeric antibody thus generated can have improved internalization capability and/or increased complement-mediated cell killing and antibody-dependent cellular cytotoxicity (ADCC). See Caron et al., *J. Exp Med.*, 176: 1191-1195 (1992) and Shopes, *J. Immunol.*, 148: 2918-2922 (1992). Homodimeric antibodies with enhanced anti-tumor activity can also be

10 prepared using heterobifunctional cross-linkers as described in Wolff et al. *Cancer Research*, 53: 2560-2565 (1993). Alternatively, an antibody can be engineered that has dual Fc regions and can thereby have enhanced complement lysis and ADCC capabilities. See Stevenson et al., *Anti-Cancer Drug Design*, 3: 219-230 (1989).

### 15 Immunoconjugates

The invention also pertains to immunoconjugates comprising an antibody conjugated to a cytotoxic agent such as a chemotherapeutic agent, toxin (*e.g.*, an enzymatically active toxin of bacterial, fungal, plant, or animal origin, or fragments thereof), or a radioactive isotope (*i.e.*, a radioconjugate).

20 Chemotherapeutic agents useful in the generation of such immunoconjugates have been described above. Enzymatically active toxins and fragments thereof that can be used include diphtheria A chain, nonbinding active fragments of diphtheria toxin, exotoxin A chain (from *Pseudomonas aeruginosa*), ricin A chain, abrin A chain, modeccin A chain, alpha-sarcin, Aleurites fordii proteins, dianthin proteins, *Phytolaca americana* proteins

25 (PAPI, PAPII, and PAP-S), momordica charantia inhibitor, curcin, crotin, sapaonaria officinalis inhibitor, gelonin, mitogellin, restrictocin, phenomycin, enomycin, and the tricothecenes. A variety of radionuclides are available for the production of radioconjugated antibodies. Examples include  $^{212}\text{Bi}$ ,  $^{131}\text{I}$ ,  $^{131}\text{In}$ ,  $^{90}\text{Y}$ , and  $^{186}\text{Re}$ .

Conjugates of the antibody and cytotoxic agent are made using a variety of

30 bifunctional protein-coupling agents such as N-succinimidyl-3-(2-pyridyldithiol) propionate (SPDP), iminothiolane (IT), bifunctional derivatives of imidoesters (such as dimethyl adipimidate HCL), active esters (such as disuccinimidyl suberate), aldehydes (such as glutaraldehyde), bis-azido compounds (such as bis (p-azidobenzoyl)

hexanediamine), bis-diazonium derivatives (such as bis-(p-diazoniumbenzoyl)-ethylenediamine), diisocyanates (such as tolyene 2,6-diisocyanate), and bis-active fluorine compounds (such as 1,5-difluoro-2,4-dinitrobenzene). For example, a ricin immunotoxin can be prepared as described in Vitetta et al., Science, 238: 1098 (1987). Carbon-14-labeled 1-isothiocyanatobenzyl-3-methyldiethylene triaminepentaacetic acid (MX-DTPA) is an exemplary chelating agent for conjugation of radionucleotide to the antibody. See WO94/11026.

In another embodiment, the antibody can be conjugated to a "receptor" (such as streptavidin) for utilization in tumor pretargeting wherein the antibody-receptor conjugate is administered to the patient, followed by removal of unbound conjugate from the circulation using a clearing agent and then administration of a "ligand" (e.g., avidin) that is in turn conjugated to a cytotoxic agent.

#### 15 **Immunoliposomes**

The antibodies disclosed herein can also be formulated as immunoliposomes. Liposomes containing the antibody are prepared by methods known in the art, such as described in Epstein et al., Proc. Natl. Acad. Sci. USA, 82: 3688 (1985); Hwang et al., Proc. Natl. Acad. Sci. USA, 77: 4030 (1980); and U.S. Pat. Nos. 4,485,045 and 4,544,545. Liposomes with enhanced circulation time are disclosed in U.S. Patent No. 5,013,556.

Particularly useful liposomes can be generated by the reverse-phase evaporation method with a lipid composition comprising phosphatidylcholine, cholesterol, and PEG-derivatized phosphatidylethanolamine (PEG-PE). Liposomes are extruded through filters of defined pore size to yield liposomes with the desired diameter. Fab' fragments of the antibody of the present invention can be conjugated to the liposomes as described in Martin et al., J. Biol. Chem., 257: 286-288 (1982) via a disulfide-interchange reaction. A chemotherapeutic agent (such as Doxorubicin) is optionally contained within the liposome. See Gabizon *et al.*, J. National Cancer Inst., 81(19): 1484 (1989).

#### 30 **Diagnostic Applications of Antibodies Directed Against the Proteins of the Invention**

In one embodiment, methods for the screening of antibodies that possess the desired specificity include, but are not limited to, enzyme linked immunosorbent assay (ELISA) and other immunologically mediated techniques known within the art. In a specific

embodiment, selection of antibodies that are specific to a particular domain of an NOVX protein is facilitated by generation of hybridomas that bind to the fragment of an NOVX protein possessing such a domain. Thus, antibodies that are specific for a desired domain within an NOVX protein, or derivatives, fragments, analogs or homologs thereof, are also  
5 provided herein.

Antibodies directed against a NOVX protein of the invention may be used in methods known within the art relating to the localization and/or quantitation of a NOVX protein (*e.g.*, for use in measuring levels of the NOVX protein within appropriate physiological samples, for use in diagnostic methods, for use in imaging the protein, and  
10 the like). In a given embodiment, antibodies specific to a NOVX protein, or derivative, fragment, analog or homolog thereof, that contain the antibody derived antigen binding domain, are utilized as pharmacologically active compounds (referred to hereinafter as "Therapeutics").

An antibody specific for a NOVX protein of the invention (*e.g.*, a monoclonal  
15 antibody or a polyclonal antibody) can be used to isolate a NOVX polypeptide by standard techniques, such as immunoaffinity, chromatography or immunoprecipitation. An antibody to a NOVX polypeptide can facilitate the purification of a natural NOVX antigen from cells, or of a recombinantly produced NOVX antigen expressed in host cells. Moreover, such an anti-NOVX antibody can be used to detect the antigenic NOVX protein (*e.g.*, in a  
20 cellular lysate or cell supernatant) in order to evaluate the abundance and pattern of expression of the antigenic NOVX protein. Antibodies directed against a NOVX protein can be used diagnostically to monitor protein levels in tissue as part of a clinical testing procedure, *e.g.*, to, for example, determine the efficacy of a given treatment regimen. Detection can be facilitated by coupling (*i.e.*, physically linking) the antibody to a  
25 detectable substance. Examples of detectable substances include various enzymes, prosthetic groups, fluorescent materials, luminescent materials, bioluminescent materials, and radioactive materials. Examples of suitable enzymes include horseradish peroxidase, alkaline phosphatase,  $\beta$ -galactosidase, or acetylcholinesterase; examples of suitable prosthetic group complexes include streptavidin/biotin and avidin/biotin; examples of  
30 suitable fluorescent materials include umbelliferone, fluorescein, fluorescein isothiocyanate, rhodamine, dichlorotriazinylamine fluorescein, dansyl chloride or phycoerythrin; an example of a luminescent material includes luminol; examples of

bioluminescent materials include luciferase, luciferin, and aequorin, and examples of suitable radioactive material include  $^{125}\text{I}$ ,  $^{131}\text{I}$ ,  $^{35}\text{S}$  or  $^3\text{H}$ .

### Antibody Therapeutics

5           Antibodies of the invention, including polyclonal, monoclonal, humanized and fully human antibodies, may used as therapeutic agents. Such agents will generally be employed to treat or prevent a disease or pathology in a subject. An antibody preparation, preferably one having high specificity and high affinity for its target antigen, is administered to the subject and will generally have an effect due to its binding with the target. Such an effect  
10   may be one of two kinds, depending on the specific nature of the interaction between the given antibody molecule and the target antigen in question. In the first instance, administration of the antibody may abrogate or inhibit the binding of the target with an endogenous ligand to which it naturally binds. In this case, the antibody binds to the target and masks a binding site of the naturally occurring ligand, wherein the ligand serves as an  
15   effector molecule. Thus the receptor mediates a signal transduction pathway for which ligand is responsible.

          Alternatively, the effect may be one in which the antibody elicits a physiological result by virtue of binding to an effector binding site on the target molecule. In this case the target, a receptor having an endogenous ligand which may be absent or defective in the  
20   disease or pathology, binds the antibody as a surrogate effector ligand, initiating a receptor-based signal transduction event by the receptor.

          A therapeutically effective amount of an antibody of the invention relates generally to the amount needed to achieve a therapeutic objective. As noted above, this may be a binding interaction between the antibody and its target antigen that, in certain cases,  
25   interferes with the functioning of the target, and in other cases, promotes a physiological response. The amount required to be administered will furthermore depend on the binding affinity of the antibody for its specific antigen, and will also depend on the rate at which an administered antibody is depleted from the free volume other subject to which it is administered. Common ranges for therapeutically effective dosing of an antibody or  
30   antibody fragment of the invention may be, by way of nonlimiting example, from about 0.1 mg/kg body weight to about 50 mg/kg body weight. Common dosing frequencies may range, for example, from twice daily to once a week.

### Pharmaceutical Compositions of Antibodies

Antibodies specifically binding a protein of the invention, as well as other molecules identified by the screening assays disclosed herein, can be administered for the treatment of various disorders in the form of pharmaceutical compositions. Principles and considerations involved in preparing such compositions, as well as guidance in the choice of components are provided, for example, in Remington : The Science And Practice Of Pharmacy 19th ed. (Alfonso R. Gennaro, et al., editors) Mack Pub. Co., Easton, Pa. : 1995; Drug Absorption Enhancement : Concepts, Possibilities, Limitations, And Trends, Harwood Academic Publishers, Langhorne, Pa., 1994; and Peptide And Protein Drug Delivery (Advances In Parenteral Sciences, Vol. 4), 1991, M. Dekker, New York.

If the antigenic protein is intracellular and whole antibodies are used as inhibitors, internalizing antibodies are preferred. However, liposomes can also be used to deliver the antibody, or an antibody fragment, into cells. Where antibody fragments are used, the smallest inhibitory fragment that specifically binds to the binding domain of the target protein is preferred. For example, based upon the variable-region sequences of an antibody, peptide molecules can be designed that retain the ability to bind the target protein sequence. Such peptides can be synthesized chemically and/or produced by recombinant DNA technology. See, e.g., Marasco et al., Proc. Natl. Acad. Sci. USA, 90: 7889-7893 (1993). The formulation herein can also contain more than one active compound as necessary for the particular indication being treated, preferably those with complementary activities that do not adversely affect each other. Alternatively, or in addition, the composition can comprise an agent that enhances its function, such as, for example, a cytotoxic agent, cytokine, chemotherapeutic agent, or growth-inhibitory agent. Such molecules are suitably present in combination in amounts that are effective for the purpose intended.

The active ingredients can also be entrapped in microcapsules prepared, for example, by coacervation techniques or by interfacial polymerization, for example, hydroxymethylcellulose or gelatin-microcapsules and poly-(methylmethacrylate) microcapsules, respectively, in colloidal drug delivery systems (for example, liposomes, albumin microspheres, microemulsions, nano-particles, and nanocapsules) or in macroemulsions.

The formulations to be used for in vivo administration must be sterile. This is readily accomplished by filtration through sterile filtration membranes.

Sustained-release preparations can be prepared. Suitable examples of sustained-release preparations include semipermeable matrices of solid hydrophobic polymers containing the antibody, which matrices are in the form of shaped articles, e.g., films, or microcapsules. Examples of sustained-release matrices include polyesters, hydrogels (for example, poly(2-hydroxyethyl-methacrylate), or poly(vinylalcohol)), polylactides (U.S. Pat. No. 3,773,919), copolymers of L-glutamic acid and  $\gamma$  ethyl-L-glutamate, non-degradable ethylene-vinyl acetate, degradable lactic acid-glycolic acid copolymers such as the LUPRON DEPOT<sup>TM</sup> (injectable microspheres composed of lactic acid-glycolic acid copolymer and leuprolide acetate), and poly-D-(-)-3-hydroxybutyric acid. While polymers such as ethylene-vinyl acetate and lactic acid-glycolic acid enable release of molecules for over 100 days, certain hydrogels release proteins for shorter time periods.

#### ELISA Assay

An agent for detecting an analyte protein is an antibody capable of binding to an analyte protein, preferably an antibody with a detectable label. Antibodies can be polyclonal, or more preferably, monoclonal. An intact antibody, or a fragment thereof (e.g., F<sub>ab</sub> or F<sub>(ab)2</sub>) can be used. The term "labeled", with regard to the probe or antibody, is intended to encompass direct labeling of the probe or antibody by coupling (*i.e.*, physically linking) a detectable substance to the probe or antibody, as well as indirect labeling of the probe or antibody by reactivity with another reagent that is directly labeled. Examples of indirect labeling include detection of a primary antibody using a fluorescently-labeled secondary antibody and end-labeling of a DNA probe with biotin such that it can be detected with fluorescently-labeled streptavidin. The term "biological sample" is intended to include tissues, cells and biological fluids isolated from a subject, as well as tissues, cells and fluids present within a subject. Included within the usage of the term "biological sample", therefore, is blood and a fraction or component of blood including blood serum, blood plasma, or lymph. That is, the detection method of the invention can be used to detect an analyte mRNA, protein, or genomic DNA in a biological sample *in vitro* as well as *in vivo*. For example, *in vitro* techniques for detection of an analyte mRNA include Northern hybridizations and *in situ* hybridizations. *In vitro* techniques for detection of an analyte protein include enzyme linked immunosorbent assays (ELISAs), Western blots, immunoprecipitations, and immunofluorescence. *In vitro* techniques for detection of an

analyte genomic DNA include Southern hybridizations. Procedures for conducting immunoassays are described, for example in "ELISA: Theory and Practice: Methods in Molecular Biology", Vol. 42, J. R. Crowther (Ed.) Human Press, Totowa, NJ, 1995; "Immunoassay", E. Diamandis and T. Christopoulos, Academic Press, Inc., San Diego, CA, 1996; and "Practice and Theory of Enzyme Immunoassays", P. Tijssen, Elsevier Science Publishers, Amsterdam, 1985. Furthermore, *in vivo* techniques for detection of an analyte protein include introducing into a subject a labeled anti-analyte protein antibody. For example, the antibody can be labeled with a radioactive marker whose presence and location in a subject can be detected by standard imaging techniques.

10

### **NOVX Recombinant Expression Vectors and Host Cells**

Another aspect of the invention pertains to vectors, preferably expression vectors, containing a nucleic acid encoding a NOVX protein, or derivatives, fragments, analogs or homologs thereof. As used herein, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of vector is a "plasmid", which refers to a circular double stranded DNA loop into which additional DNA segments can be ligated. Another type of vector is a viral vector, wherein additional DNA segments can be ligated into the viral genome. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (*e.g.*, bacterial vectors having a bacterial origin of replication and episomal mammalian vectors). Other vectors (*e.g.*, non-episomal mammalian vectors) are integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors are capable of directing the expression of genes to which they are operatively-linked. Such vectors are referred to herein as "expression vectors". In general, expression vectors of utility in recombinant DNA techniques are often in the form of plasmids. In the present specification, "plasmid" and "vector" can be used interchangeably as the plasmid is the most commonly used form of vector. However, the invention is intended to include such other forms of expression vectors, such as viral vectors (*e.g.*, replication defective retroviruses, adenoviruses and adeno-associated viruses), which serve equivalent functions.

30

The recombinant expression vectors of the invention comprise a nucleic acid of the invention in a form suitable for expression of the nucleic acid in a host cell, which means that the recombinant expression vectors include one or more regulatory sequences, selected

on the basis of the host cells to be used for expression, that is operatively-linked to the nucleic acid sequence to be expressed. Within a recombinant expression vector, "operably-linked" is intended to mean that the nucleotide sequence of interest is linked to the regulatory sequence(s) in a manner that allows for expression of the nucleotide sequence (*e.g.*, in an *in vitro* transcription/translation system or in a host cell when the vector is introduced into the host cell).

The term "regulatory sequence" is intended to include promoters, enhancers and other expression control elements (*e.g.*, polyadenylation signals). Such regulatory sequences are described, for example, in Goeddel, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990). Regulatory sequences include those that direct constitutive expression of a nucleotide sequence in many types of host cell and those that direct expression of the nucleotide sequence only in certain host cells (*e.g.*, tissue-specific regulatory sequences). It will be appreciated by those skilled in the art that the design of the expression vector can depend on such factors as the choice of the host cell to be transformed, the level of expression of protein desired, *etc.* The expression vectors of the invention can be introduced into host cells to thereby produce proteins or peptides, including fusion proteins or peptides, encoded by nucleic acids as described herein (*e.g.*, NOVX proteins, mutant forms of NOVX proteins, fusion proteins, *etc.*).

The recombinant expression vectors of the invention can be designed for expression of NOVX proteins in prokaryotic or eukaryotic cells. For example, NOVX proteins can be expressed in bacterial cells such as *Escherichia coli*, insect cells (using baculovirus expression vectors) yeast cells or mammalian cells. Suitable host cells are discussed further in Goeddel, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990). Alternatively, the recombinant expression vector can be transcribed and translated *in vitro*, for example using T7 promoter regulatory sequences and T7 polymerase.

Expression of proteins in prokaryotes is most often carried out in *Escherichia coli* with vectors containing constitutive or inducible promoters directing the expression of either fusion or non-fusion proteins. Fusion vectors add a number of amino acids to a protein encoded therein, usually to the amino terminus of the recombinant protein. Such fusion vectors typically serve three purposes: (i) to increase expression of recombinant protein; (ii) to increase the solubility of the recombinant protein; and (iii) to aid in the

purification of the recombinant protein by acting as a ligand in affinity purification. Often, in fusion expression vectors, a proteolytic cleavage site is introduced at the junction of the fusion moiety and the recombinant protein to enable separation of the recombinant protein from the fusion moiety subsequent to purification of the fusion protein. Such enzymes, and  
5 their cognate recognition sequences, include Factor Xa, thrombin and enterokinase.

Typical fusion expression vectors include pGEX (Pharmacia Biotech Inc; Smith and Johnson, 1988. *Gene* 67: 31-40), pMAL (New England Biolabs, Beverly, Mass.) and pRIT5 (Pharmacia, Piscataway, N.J.) that fuse glutathione S-transferase (GST), maltose E binding protein, or protein A, respectively, to the target recombinant protein.

10 Examples of suitable inducible non-fusion *E. coli* expression vectors include pTrc (Amrann *et al.*, (1988) *Gene* 69:301-315) and pET 11d (Studier *et al.*, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990) 60-89).

One strategy to maximize recombinant protein expression in *E. coli* is to express the  
15 protein in a host bacteria with an impaired capacity to proteolytically cleave the recombinant protein. *See, e.g.*, Gottesman, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990) 119-128. Another strategy is to alter the nucleic acid sequence of the nucleic acid to be inserted into an expression vector so that the individual codons for each amino acid are those preferentially utilized in *E. coli*  
20 (*see, e.g.*, Wada, *et al.*, 1992. *Nucl. Acids Res.* 20: 2111-2118). Such alteration of nucleic acid sequences of the invention can be carried out by standard DNA synthesis techniques.

In another embodiment, the NOVX expression vector is a yeast expression vector. Examples of vectors for expression in yeast *Saccharomyces cerevisiae* include pYepSec1 (Baldari, *et al.*, 1987. *EMBO J.* 6: 229-234), pMFa (Kurjan and Herskowitz, 1982. *Cell* 30:  
25 933-943), pJRY88 (Schultz *et al.*, 1987. *Gene* 54: 113-123), pYES2 (Invitrogen Corporation, San Diego, Calif.), and picZ (InVitrogen Corp, San Diego, Calif.).

Alternatively, NOVX can be expressed in insect cells using baculovirus expression vectors. Baculovirus vectors available for expression of proteins in cultured insect cells (*e.g.*, SF9 cells) include the pAc series (Smith, *et al.*, 1983. *Mol. Cell. Biol.* 3: 2156-2165)  
30 and the pVL series (Lucklow and Summers, 1989. *Virology* 170: 31-39).

In yet another embodiment, a nucleic acid of the invention is expressed in mammalian cells using a mammalian expression vector. Examples of mammalian expression vectors include pCDM8 (Seed, 1987. *Nature* 329: 840) and pMT2PC (Kaufman,

*et al.*, 1987. *EMBO J.* 6: 187-195). When used in mammalian cells, the expression vector's control functions are often provided by viral regulatory elements. For example, commonly used promoters are derived from polyoma, adenovirus 2, cytomegalovirus, and simian virus 40. For other suitable expression systems for both prokaryotic and eukaryotic cells see, 5 *e.g.*, Chapters 16 and 17 of Sambrook, *et al.*, MOLECULAR CLONING: A LABORATORY MANUAL. 2nd ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989.

In another embodiment, the recombinant mammalian expression vector is capable of directing expression of the nucleic acid preferentially in a particular cell type (*e.g.*, 10 tissue-specific regulatory elements are used to express the nucleic acid). Tissue-specific regulatory elements are known in the art. Non-limiting examples of suitable tissue-specific promoters include the albumin promoter (liver-specific; Pinkert, *et al.*, 1987. *Genes Dev.* 1: 268-277), lymphoid-specific promoters (Calame and Eaton, 1988. *Adv. Immunol.* 43: 235-275), in particular promoters of T cell receptors (Winoto and Baltimore, 1989. *EMBO J.* 8: 729-733) and immunoglobulins (Banerji, *et al.*, 1983. *Cell* 33: 729-740; Queen and Baltimore, 1983. *Cell* 33: 741-748), neuron-specific promoters (*e.g.*, the neurofilament promoter; Byrne and Ruddle, 1989. *Proc. Natl. Acad. Sci. USA* 86: 5473-5477), pancreas-specific promoters (Edlund, *et al.*, 1985. *Science* 230: 912-916), and mammary gland-specific promoters (*e.g.*, milk whey promoter; U.S. Pat. No. 4,873,316 and European 20 Application Publication No. 264,166). Developmentally-regulated promoters are also encompassed, *e.g.*, the murine hox promoters (Kessel and Gruss, 1990. *Science* 249: 374-379) and the  $\alpha$ -fetoprotein promoter (Campes and Tilghman, 1989. *Genes Dev.* 3: 537-546).

The invention further provides a recombinant expression vector comprising a DNA 25 molecule of the invention cloned into the expression vector in an antisense orientation. That is, the DNA molecule is operatively-linked to a regulatory sequence in a manner that allows for expression (by transcription of the DNA molecule) of an RNA molecule that is antisense to NOVX mRNA. Regulatory sequences operatively linked to a nucleic acid cloned in the antisense orientation can be chosen that direct the continuous expression of 30 the antisense RNA molecule in a variety of cell types, for instance viral promoters and/or enhancers, or regulatory sequences can be chosen that direct constitutive, tissue specific or cell type specific expression of antisense RNA. The antisense expression vector can be in the form of a recombinant plasmid, phagemid or attenuated virus in which antisense nucleic

acids are produced under the control of a high efficiency regulatory region, the activity of which can be determined by the cell type into which the vector is introduced. For a discussion of the regulation of gene expression using antisense genes *see, e.g.,* Weintraub, *et al.*, "Antisense RNA as a molecular tool for genetic analysis," *Reviews-Trends in*  
5 *Genetics*, Vol. 1(1) 1986.

Another aspect of the invention pertains to host cells into which a recombinant expression vector of the invention has been introduced. The terms "host cell" and "recombinant host cell" are used interchangeably herein. It is understood that such terms refer not only to the particular subject cell but also to the progeny or potential progeny of  
10 such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term as used herein.

A host cell can be any prokaryotic or eukaryotic cell. For example, NOVX protein can be expressed in bacterial cells such as *E. coli*, insect cells, yeast or mammalian cells  
15 (such as Chinese hamster ovary cells (CHO) or COS cells). Other suitable host cells are known to those skilled in the art.

Vector DNA can be introduced into prokaryotic or eukaryotic cells via conventional transformation or transfection techniques. As used herein, the terms "transformation" and "transfection" are intended to refer to a variety of art-recognized techniques for introducing  
20 foreign nucleic acid (*e.g.*, DNA) into a host cell, including calcium phosphate or calcium chloride co-precipitation, DEAE-dextran-mediated transfection, lipofection, or electroporation. Suitable methods for transforming or transfecting host cells can be found in Sambrook, *et al.* (MOLECULAR CLONING: A LABORATORY MANUAL, 2nd ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor,  
25 N.Y., 1989), and other laboratory manuals.

For stable transfection of mammalian cells, it is known that, depending upon the expression vector and transfection technique used, only a small fraction of cells may integrate the foreign DNA into their genome. In order to identify and select these integrants, a gene that encodes a selectable marker (*e.g.*, resistance to antibiotics) is  
30 generally introduced into the host cells along with the gene of interest. Various selectable markers include those that confer resistance to drugs, such as G418, hygromycin and methotrexate. Nucleic acid encoding a selectable marker can be introduced into a host cell on the same vector as that encoding NOVX or can be introduced on a separate vector.

Cells stably transfected with the introduced nucleic acid can be identified by drug selection (*e.g.*, cells that have incorporated the selectable marker gene will survive, while the other cells die).

5 A host cell of the invention, such as a prokaryotic or eukaryotic host cell in culture, can be used to produce (*i.e.*, express) NOVX protein. Accordingly, the invention further provides methods for producing NOVX protein using the host cells of the invention. In one embodiment, the method comprises culturing the host cell of invention (into which a recombinant expression vector encoding NOVX protein has been introduced) in a suitable medium such that NOVX protein is produced. In another embodiment, the method further  
10 comprises isolating NOVX protein from the medium or the host cell.

#### **Transgenic NOVX Animals**

The host cells of the invention can also be used to produce non-human transgenic animals. For example, in one embodiment, a host cell of the invention is a fertilized oocyte  
15 or an embryonic stem cell into which NOVX protein-coding sequences have been introduced. Such host cells can then be used to create non-human transgenic animals in which exogenous NOVX sequences have been introduced into their genome or homologous recombinant animals in which endogenous NOVX sequences have been altered. Such animals are useful for studying the function and/or activity of NOVX protein  
20 and for identifying and/or evaluating modulators of NOVX protein activity. As used herein, a "transgenic animal" is a non-human animal, preferably a mammal, more preferably a rodent such as a rat or mouse, in which one or more of the cells of the animal includes a transgene. Other examples of transgenic animals include non-human primates, sheep, dogs, cows, goats, chickens, amphibians, *etc.* A transgene is exogenous DNA that is  
25 integrated into the genome of a cell from which a transgenic animal develops and that remains in the genome of the mature animal, thereby directing the expression of an encoded gene product in one or more cell types or tissues of the transgenic animal. As used herein, a "homologous recombinant animal" is a non-human animal, preferably a mammal, more preferably a mouse, in which an endogenous NOVX gene has been altered by  
30 homologous recombination between the endogenous gene and an exogenous DNA molecule introduced into a cell of the animal, *e.g.*, an embryonic cell of the animal, prior to development of the animal.

A transgenic animal of the invention can be created by introducing NOVX-encoding nucleic acid into the male pronuclei of a fertilized oocyte (*e.g.*, by microinjection, retroviral infection) and allowing the oocyte to develop in a pseudopregnant female foster animal. The human NOVX cDNA sequences, *i.e.*, any one of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 226, can be introduced as a transgene into the genome of a non-human animal. Alternatively, a non-human homologue of the human NOVX gene, such as a mouse NOVX gene, can be isolated based on hybridization to the human NOVX cDNA (described further *supra*) and used as a transgene. Intronic sequences and polyadenylation signals can also be included in the transgene to increase the efficiency of expression of the transgene. A tissue-specific regulatory sequence(s) can be operably-linked to the NOVX transgene to direct expression of NOVX protein to particular cells. Methods for generating transgenic animals via embryo manipulation and microinjection, particularly animals such as mice, have become conventional in the art and are described, for example, in U.S. Patent Nos. 4,736,866; 4,870,009; and 4,873,191; and Hogan, 1986. In: MANIPULATING THE MOUSE EMBRYO, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. Similar methods are used for production of other transgenic animals. A transgenic founder animal can be identified based upon the presence of the NOVX transgene in its genome and/or expression of NOVX mRNA in tissues or cells of the animals. A transgenic founder animal can then be used to breed additional animals carrying the transgene. Moreover, transgenic animals carrying a transgene-encoding NOVX protein can further be bred to other transgenic animals carrying other transgenes.

To create a homologous recombinant animal, a vector is prepared which contains at least a portion of a NOVX gene into which a deletion, addition or substitution has been introduced to thereby alter, *e.g.*, functionally disrupt, the NOVX gene. The NOVX gene can be a human gene (*e.g.*, the cDNA of any one of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 226), but more preferably, is a non-human homologue of a human NOVX gene. For example, a mouse homologue of human NOVX gene of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 226, can be used to construct a homologous recombination vector suitable for altering an endogenous NOVX gene in the mouse genome. In one embodiment, the vector is designed such that, upon homologous recombination, the endogenous NOVX gene is functionally disrupted (*i.e.*, no longer encodes a functional protein; also referred to as a "knock out" vector).

Alternatively, the vector can be designed such that, upon homologous recombination, the endogenous NOVX gene is mutated or otherwise altered but still encodes functional protein (e.g., the upstream regulatory region can be altered to thereby alter the expression of the endogenous NOVX protein). In the homologous recombination vector, the altered portion of the NOVX gene is flanked at its 5'- and 3'-termini by additional nucleic acid of the NOVX gene to allow for homologous recombination to occur between the exogenous NOVX gene carried by the vector and an endogenous NOVX gene in an embryonic stem cell. The additional flanking NOVX nucleic acid is of sufficient length for successful homologous recombination with the endogenous gene. Typically, several kilobases of flanking DNA (both at the 5'- and 3'-termini) are included in the vector. *See, e.g., Thomas, et al., 1987. Cell 51: 503 for a description of homologous recombination vectors. The vector is then introduced into an embryonic stem cell line (e.g., by electroporation) and cells in which the introduced NOVX gene has homologously-recombined with the endogenous NOVX gene are selected. See, e.g., Li, et al., 1992. Cell 69: 915.*

The selected cells are then injected into a blastocyst of an animal (e.g., a mouse) to form aggregation chimeras. *See, e.g., Bradley, 1987. In: TERATOCARCINOMAS AND EMBRYONIC STEM CELLS: A PRACTICAL APPROACH, Robertson, ed. IRL, Oxford, pp. 113-152. A chimeric embryo can then be implanted into a suitable pseudopregnant female foster animal and the embryo brought to term. Progeny harboring the homologously-recombined DNA in their germ cells can be used to breed animals in which all cells of the animal contain the homologously-recombined DNA by germline transmission of the transgene. Methods for constructing homologous recombination vectors and homologous recombinant animals are described further in Bradley, 1991. Curr. Opin. Biotechnol. 2: 823-829; PCT International Publication Nos.: WO 90/11354; WO 91/01140; WO 92/0968; and WO 93/04169.*

In another embodiment, transgenic non-humans animals can be produced that contain selected systems that allow for regulated expression of the transgene. One example of such a system is the cre/loxP recombinase system of bacteriophage P1. For a description of the cre/loxP recombinase system, *See, e.g., Lakso, et al., 1992. Proc. Natl. Acad. Sci. USA 89: 6232-6236. Another example of a recombinase system is the FLP recombinase system of Saccharomyces cerevisiae. See, O'Gorman, et al., 1991. Science 251:1351-1355. If a cre/loxP recombinase system is used to regulate expression of the transgene, animals*

containing transgenes encoding both the Cre recombinase and a selected protein are required. Such animals can be provided through the construction of "double" transgenic animals, *e.g.*, by mating two transgenic animals, one containing a transgene encoding a selected protein and the other containing a transgene encoding a recombinase.

- 5           Clones of the non-human transgenic animals described herein can also be produced according to the methods described in Wilmut, *et al.*, 1997. *Nature* 385: 810-813. In brief, a cell (*e.g.*, a somatic cell) from the transgenic animal can be isolated and induced to exit the growth cycle and enter G<sub>0</sub> phase. The quiescent cell can then be fused, *e.g.*, through the use of electrical pulses, to an enucleated oocyte from an animal of the same species from  
10   which the quiescent cell is isolated. The reconstructed oocyte is then cultured such that it develops to morula or blastocyte and then transferred to pseudopregnant female foster animal. The offspring borne of this female foster animal will be a clone of the animal from which the cell (*e.g.*, the somatic cell) is isolated.

#### 15. **Pharmaceutical Compositions**

- The NOVX nucleic acid molecules, NOVX proteins, and anti-NOVX antibodies (also referred to herein as "active compounds") of the invention, and derivatives, fragments, analogs and homologs thereof, can be incorporated into pharmaceutical compositions suitable for administration. Such compositions typically comprise the nucleic acid  
20   molecule, protein, or antibody and a pharmaceutically acceptable carrier. As used herein, "pharmaceutically acceptable carrier" is intended to include any and all solvents, dispersion media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents, and the like, compatible with pharmaceutical administration. Suitable carriers are described in the most recent edition of Remington's Pharmaceutical Sciences, a standard  
25   reference text in the field, which is incorporated herein by reference. Preferred examples of such carriers or diluents include, but are not limited to, water, saline, finger's solutions, dextrose solution, and 5% human serum albumin. Liposomes and non-aqueous vehicles such as fixed oils may also be used. The use of such media and agents for  
30   pharmaceutically active substances is well known in the art. Except insofar as any conventional media or agent is incompatible with the active compound, use thereof in the compositions is contemplated. Supplementary active compounds can also be incorporated into the compositions.

A pharmaceutical composition of the invention is formulated to be compatible with its intended route of administration. Examples of routes of administration include parenteral, *e.g.*, intravenous, intradermal, subcutaneous, oral (*e.g.*, inhalation), transdermal (*i.e.*, topical), transmucosal, and rectal administration. Solutions or suspensions used for parenteral, intradermal, or subcutaneous application can include the following components: a sterile diluent such as water for injection, saline solution, fixed oils, polyethylene glycols, glycerine, propylene glycol or other synthetic solvents; antibacterial agents such as benzyl alcohol or methyl parabens; antioxidants such as ascorbic acid or sodium bisulfite; chelating agents such as ethylenediaminetetraacetic acid (EDTA); buffers such as acetates, citrates or phosphates, and agents for the adjustment of tonicity such as sodium chloride or dextrose. The pH can be adjusted with acids or bases, such as hydrochloric acid or sodium hydroxide. The parenteral preparation can be enclosed in ampoules, disposable syringes or multiple dose vials made of glass or plastic.

Pharmaceutical compositions suitable for injectable use include sterile aqueous solutions (where water soluble) or dispersions and sterile powders for the extemporaneous preparation of sterile injectable solutions or dispersion. For intravenous administration, suitable carriers include physiological saline, bacteriostatic water, Cremophor EL™ (BASF, Parsippany, N.J.) or phosphate buffered saline (PBS). In all cases, the composition must be sterile and should be fluid to the extent that easy syringeability exists. It must be stable under the conditions of manufacture and storage and must be preserved against the contaminating action of microorganisms such as bacteria and fungi. The carrier can be a solvent or dispersion medium containing, for example, water, ethanol, polyol (for example, glycerol, propylene glycol, and liquid polyethylene glycol, and the like), and suitable mixtures thereof. The proper fluidity can be maintained, for example, by the use of a coating such as lecithin, by the maintenance of the required particle size in the case of dispersion and by the use of surfactants. Prevention of the action of microorganisms can be achieved by various antibacterial and antifungal agents, for example, parabens, chlorobutanol, phenol, ascorbic acid, thimerosal, and the like. In many cases, it will be preferable to include isotonic agents, for example, sugars, polyalcohols such as manitol, sorbitol, sodium chloride in the composition. Prolonged absorption of the injectable compositions can be brought about by including in the composition an agent which delays absorption, for example, aluminum monostearate and gelatin.

Sterile injectable solutions can be prepared by incorporating the active compound (e.g., a NOVX protein or anti-NOVX antibody) in the required amount in an appropriate solvent with one or a combination of ingredients enumerated above, as required, followed by filtered sterilization. Generally, dispersions are prepared by incorporating the active  
5 compound into a sterile vehicle that contains a basic dispersion medium and the required other ingredients from those enumerated above. In the case of sterile powders for the preparation of sterile injectable solutions, methods of preparation are vacuum drying and freeze-drying that yields a powder of the active ingredient plus any additional desired ingredient from a previously sterile-filtered solution thereof.

10 Oral compositions generally include an inert diluent or an edible carrier. They can be enclosed in gelatin capsules or compressed into tablets. For the purpose of oral therapeutic administration, the active compound can be incorporated with excipients and used in the form of tablets, troches, or capsules. Oral compositions can also be prepared using a fluid carrier for use as a mouthwash, wherein the compound in the fluid carrier is  
15 applied orally and swished and expectorated or swallowed. Pharmaceutically compatible binding agents, and/or adjuvant materials can be included as part of the composition. The tablets, pills, capsules, troches and the like can contain any of the following ingredients, or compounds of a similar nature: a binder such as microcrystalline cellulose, gum tragacanth or gelatin; an excipient such as starch or lactose, a disintegrating agent such as alginic acid, Primogel, or corn starch; a lubricant such as magnesium stearate or Sterotes; a glidant such  
20 as colloidal silicon dioxide; a sweetening agent such as sucrose or saccharin; or a flavoring agent such as peppermint, methyl salicylate, or orange flavoring.

For administration by inhalation, the compounds are delivered in the form of an aerosol spray from pressured container or dispenser which contains a suitable propellant,  
25 e.g., a gas such as carbon dioxide, or a nebulizer.

Systemic administration can also be by transmucosal or transdermal means. For transmucosal or transdermal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art, and include, for example, for transmucosal administration, detergents, bile salts, and fusidic  
30 acid derivatives. Transmucosal administration can be accomplished through the use of nasal sprays or suppositories. For transdermal administration, the active compounds are formulated into ointments, salves, gels, or creams as generally known in the art.

The compounds can also be prepared in the form of suppositories (*e.g.*, with conventional suppository bases such as cocoa butter and other glycerides) or retention enemas for rectal delivery.

In one embodiment, the active compounds are prepared with carriers that will  
5 protect the compound against rapid elimination from the body, such as a controlled release formulation, including implants and microencapsulated delivery systems. Biodegradable, biocompatible polymers can be used, such as ethylene vinyl acetate, polyanhydrides, polyglycolic acid, collagen, polyorthoesters, and polylactic acid. Methods for preparation of such formulations will be apparent to those skilled in the art. The materials can also be  
10 obtained commercially from Alza Corporation and Nova Pharmaceuticals, Inc. Liposomal suspensions (including liposomes targeted to infected cells with monoclonal antibodies to viral antigens) can also be used as pharmaceutically acceptable carriers. These can be prepared according to methods known to those skilled in the art, for example, as described in U.S. Patent No. 4,522,811.

15 It is especially advantageous to formulate oral or parenteral compositions in dosage unit form for ease of administration and uniformity of dosage. Dosage unit form as used herein refers to physically discrete units suited as unitary dosages for the subject to be treated; each unit containing a predetermined quantity of active compound calculated to produce the desired therapeutic effect in association with the required pharmaceutical  
20 carrier. The specification for the dosage unit forms of the invention are dictated by and directly dependent on the unique characteristics of the active compound and the particular therapeutic effect to be achieved, and the limitations inherent in the art of compounding such an active compound for the treatment of individuals.

The nucleic acid molecules of the invention can be inserted into vectors and used as  
25 gene therapy vectors. Gene therapy vectors can be delivered to a subject by, for example, intravenous injection, local administration (*see, e.g.*, U.S. Patent No. 5,328,470) or by stereotactic injection (*see, e.g.*, Chen, *et al.*, 1994. *Proc. Natl. Acad. Sci. USA* 91: 3054-3057). The pharmaceutical preparation of the gene therapy vector can include the gene therapy vector in an acceptable diluent, or can comprise a slow release matrix in  
30 which the gene delivery vehicle is imbedded. Alternatively, where the complete gene delivery vector can be produced intact from recombinant cells, *e.g.*, retroviral vectors, the pharmaceutical preparation can include one or more cells that produce the gene delivery system.

The pharmaceutical compositions can be included in a container, pack, or dispenser together with instructions for administration.

### Screening and Detection Methods

5. The isolated nucleic acid molecules of the invention can be used to express NOVX protein (*e.g.*, via a recombinant expression vector in a host cell in gene therapy applications), to detect NOVX mRNA (*e.g.*, in a biological sample) or a genetic lesion in a NOVX gene, and to modulate NOVX activity, as described further, below. In addition, the NOVX proteins can be used to screen drugs or compounds that modulate the NOVX
- 10 protein activity or expression as well as to treat disorders characterized by insufficient or excessive production of NOVX protein or production of NOVX protein forms that have decreased or aberrant activity compared to NOVX wild-type protein (*e.g.*; diabetes (regulates insulin release); obesity (binds and transport lipids); metabolic disturbances associated with obesity, the metabolic syndrome X as well as anorexia and wasting
- 15 disorders associated with chronic diseases and various cancers, and infectious disease (possesses anti-microbial activity) and the various dyslipidemias. In addition, the anti-NOVX antibodies of the invention can be used to detect and isolate NOVX proteins and modulate NOVX activity. In yet a further aspect, the invention can be used in methods to influence appetite, absorption of nutrients and the disposition of metabolic substrates in
- 20 both a positive and negative fashion.

The invention further pertains to novel agents identified by the screening assays described herein and uses thereof for treatments as described, *supra*.

### Screening Assays

- 25 The invention provides a method (also referred to herein as a "screening assay") for identifying modulators, *i.e.*, candidate or test compounds or agents (*e.g.*, peptides, peptidomimetics, small molecules or other drugs) that bind to NOVX proteins or have a stimulatory or inhibitory effect on, *e.g.*, NOVX protein expression or NOVX protein activity. The invention also includes compounds identified in the screening assays
- 30 described herein.

In one embodiment, the invention provides assays for screening candidate or test compounds which bind to or modulate the activity of the membrane-bound form of a NOVX protein or polypeptide or biologically-active portion thereof. The test compounds

of the invention can be obtained using any of the numerous approaches in combinatorial library methods known in the art, including: biological libraries; spatially addressable parallel solid phase or solution phase libraries; synthetic library methods requiring deconvolution; the "one-bead one-compound" library method; and synthetic library methods using affinity chromatography selection. The biological library approach is limited to peptide libraries, while the other four approaches are applicable to peptide, non-peptide oligomer or small molecule libraries of compounds. *See, e.g., Lam, 1997. Anticancer Drug Design 12: 145.*

A "small molecule" as used herein, is meant to refer to a composition that has a molecular weight of less than about 5 kD and most preferably less than about 4 kD. Small molecules can be, *e.g.*, nucleic acids, peptides, polypeptides, peptidomimetics, carbohydrates, lipids or other organic or inorganic molecules. Libraries of chemical and/or biological mixtures, such as fungal, bacterial, or algal extracts, are known in the art and can be screened with any of the assays of the invention.

Examples of methods for the synthesis of molecular libraries can be found in the art, for example in: DeWitt, *et al.*, 1993. *Proc. Natl. Acad. Sci. U.S.A.* 90: 6909; Erb, *et al.*, 1994. *Proc. Natl. Acad. Sci. U.S.A.* 91: 11422; Zuckermann, *et al.*, 1994. *J. Med. Chem.* 37: 2678; Cho, *et al.*, 1993. *Science* 261: 1303; Carrell, *et al.*, 1994. *Angew. Chem. Int. Ed. Engl.* 33: 2059; Carell, *et al.*, 1994. *Angew. Chem. Int. Ed. Engl.* 33: 2061; and Gallop, *et al.*, 1994. *J. Med. Chem.* 37: 1233.

Libraries of compounds may be presented in solution (*e.g.*, Houghten, 1992. *Biotechniques* 13: 412-421), or on beads (Lam, 1991. *Nature* 354: 82-84), on chips (Fodor, 1993. *Nature* 364: 555-556), bacteria (Ladner, U.S. Patent No. 5,223,409), spores (Ladner, U.S. Patent 5,233,409), plasmids (Cull, *et al.*, 1992. *Proc. Natl. Acad. Sci. USA* 89: 1865-1869) or on phage (Scott and Smith, 1990. *Science* 249: 386-390; Devlin, 1990. *Science* 249: 404-406; Cwirla, *et al.*, 1990. *Proc. Natl. Acad. Sci. U.S.A.* 87: 6378-6382; Felici, 1991. *J. Mol. Biol.* 222: 301-310; Ladner, U.S. Patent No. 5,233,409.).

In one embodiment, an assay is a cell-based assay in which a cell which expresses a membrane-bound form of NOVX protein, or a biologically-active portion thereof, on the cell surface is contacted with a test compound and the ability of the test compound to bind to a NOVX protein determined. The cell, for example, can be of mammalian origin or a yeast cell. Determining the ability of the test compound to bind to the NOVX protein can be accomplished, for example, by coupling the test compound with a radioisotope or

enzymatic label such that binding of the test compound to the NOVX protein or biologically-active portion thereof can be determined by detecting the labeled compound in a complex. For example, test compounds can be labeled with  $^{125}\text{I}$ ,  $^{35}\text{S}$ ,  $^{14}\text{C}$ , or  $^3\text{H}$ , either directly or indirectly, and the radioisotope detected by direct counting of radioemission or by scintillation counting. Alternatively, test compounds can be enzymatically-labeled with, for example, horseradish peroxidase, alkaline phosphatase, or luciferase, and the enzymatic label detected by determination of conversion of an appropriate substrate to product. In one embodiment, the assay comprises contacting a cell which expresses a membrane-bound form of NOVX protein, or a biologically-active portion thereof, on the cell surface with a known compound which binds NOVX to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with a NOVX protein, wherein determining the ability of the test compound to interact with a NOVX protein comprises determining the ability of the test compound to preferentially bind to NOVX protein or a biologically-active portion thereof as compared to the known compound.

In another embodiment, an assay is a cell-based assay comprising contacting a cell expressing a membrane-bound form of NOVX protein, or a biologically-active portion thereof, on the cell surface with a test compound and determining the ability of the test compound to modulate (*e.g.*, stimulate or inhibit) the activity of the NOVX protein or biologically-active portion thereof. Determining the ability of the test compound to modulate the activity of NOVX or a biologically-active portion thereof can be accomplished, for example, by determining the ability of the NOVX protein to bind to or interact with a NOVX target molecule. As used herein, a "target molecule" is a molecule with which a NOVX protein binds or interacts in nature, for example, a molecule on the surface of a cell which expresses a NOVX interacting protein, a molecule on the surface of a second cell, a molecule in the extracellular milieu, a molecule associated with the internal surface of a cell membrane or a cytoplasmic molecule. A NOVX target molecule can be a non-NOVX molecule or a NOVX protein or polypeptide of the invention. In one embodiment, a NOVX target molecule is a component of a signal transduction pathway that facilitates transduction of an extracellular signal (*e.g.* a signal generated by binding of a compound to a membrane-bound NOVX molecule) through the cell membrane and into the cell. The target, for example, can be a second intercellular protein that has catalytic

activity or a protein that facilitates the association of downstream signaling molecules with NOVX.

Determining the ability of the NOVX protein to bind to or interact with a NOVX target molecule can be accomplished by one of the methods described above for  
5 determining direct binding. In one embodiment, determining the ability of the NOVX protein to bind to or interact with a NOVX target molecule can be accomplished by determining the activity of the target molecule. For example, the activity of the target molecule can be determined by detecting induction of a cellular second messenger of the target (*i.e.* intracellular  $\text{Ca}^{2+}$ , diacylglycerol,  $\text{IP}_3$ , *etc.*), detecting catalytic/enzymatic  
10 activity of the target on an appropriate substrate, detecting the induction of a reporter gene (comprising a NOVX-responsive regulatory element operatively linked to a nucleic acid encoding a detectable marker, *e.g.*, luciferase), or detecting a cellular response, for example, cell survival, cellular differentiation, or cell proliferation.

In yet another embodiment, an assay of the invention is a cell-free assay comprising  
15 contacting a NOVX protein or biologically-active portion thereof with a test compound and determining the ability of the test compound to bind to the NOVX protein or biologically-active portion thereof. Binding of the test compound to the NOVX protein can be determined either directly or indirectly as described above. In one such embodiment, the assay comprises contacting the NOVX protein or biologically-active portion thereof  
20 with a known compound which binds NOVX to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with a NOVX protein, wherein determining the ability of the test compound to interact with a NOVX protein comprises determining the ability of the test compound to preferentially bind to NOVX or biologically-active portion thereof as compared to the  
25 known compound.

In still another embodiment, an assay is a cell-free assay comprising contacting NOVX protein or biologically-active portion thereof with a test compound and determining the ability of the test compound to modulate (*e.g.* stimulate or inhibit) the activity of the NOVX protein or biologically-active portion thereof. Determining the ability of the test  
30 compound to modulate the activity of NOVX can be accomplished, for example, by determining the ability of the NOVX protein to bind to a NOVX target molecule by one of the methods described above for determining direct binding. In an alternative embodiment, determining the ability of the test compound to modulate the activity of NOVX protein can

be accomplished by determining the ability of the NOVX protein further modulate a NOVX target molecule. For example, the catalytic/enzymatic activity of the target molecule on an appropriate substrate can be determined as described, *supra*.

In yet another embodiment, the cell-free assay comprises contacting the NOVX protein or biologically-active portion thereof with a known compound which binds NOVX protein to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with a NOVX protein, wherein determining the ability of the test compound to interact with a NOVX protein comprises determining the ability of the NOVX protein to preferentially bind to or modulate the activity of a NOVX target molecule.

The cell-free assays of the invention are amenable to use of both the soluble form or the membrane-bound form of NOVX protein. In the case of cell-free assays comprising the membrane-bound form of NOVX protein, it may be desirable to utilize a solubilizing agent such that the membrane-bound form of NOVX protein is maintained in solution. Examples of such solubilizing agents include non-ionic detergents such as n-octylglucoside, n-dodecylglucoside, n-dodecylmaltoside, octanoyl-N-methylglucamide, decanoyl-N-methylglucamide, Triton<sup>®</sup> X-100, Triton<sup>®</sup> X-114, Thesit<sup>®</sup>, Isotridecypoly(ethylene glycol ether)<sub>n</sub>, N-dodecyl--N,N-dimethyl-3-ammonio-1-propane sulfonate, 3-(3-cholamidopropyl) dimethylamminiol-1-propane sulfonate (CHAPS), or 3-(3-cholamidopropyl)dimethylamminiol-2-hydroxy-1-propane sulfonate (CHAPSO).

In more than one embodiment of the above assay methods of the invention, it may be desirable to immobilize either NOVX protein or its target molecule to facilitate separation of complexed from uncomplexed forms of one or both of the proteins, as well as to accommodate automation of the assay. Binding of a test compound to NOVX protein, or interaction of NOVX protein with a target molecule in the presence and absence of a candidate compound, can be accomplished in any vessel suitable for containing the reactants. Examples of such vessels include microtiter plates, test tubes, and micro-centrifuge tubes. In one embodiment, a fusion protein can be provided that adds a domain that allows one or both of the proteins to be bound to a matrix. For example, GST-NOVX fusion proteins or GST-target fusion proteins can be adsorbed onto glutathione sepharose beads (Sigma Chemical, St. Louis, MO) or glutathione derivatized microtiter plates, that are then combined with the test compound or the test compound and either the non-adsorbed target protein or NOVX protein, and the mixture is incubated under

conditions conducive to complex formation (*e.g.*, at physiological conditions for salt and pH). Following incubation, the beads or microtiter plate wells are washed to remove any unbound components, the matrix immobilized in the case of beads, complex determined either directly or indirectly, for example, as described, *supra*. Alternatively, the complexes  
5 can be dissociated from the matrix, and the level of NOVX protein binding or activity determined using standard techniques.

Other techniques for immobilizing proteins on matrices can also be used in the screening assays of the invention. For example, either the NOVX protein or its target molecule can be immobilized utilizing conjugation of biotin and streptavidin. Biotinylated  
10 NOVX protein or target molecules can be prepared from biotin-NHS (N-hydroxy-succinimide) using techniques well-known within the art (*e.g.*, biotinylation kit, Pierce Chemicals, Rockford, Ill.), and immobilized in the wells of streptavidin-coated 96 well plates (Pierce Chemical). Alternatively, antibodies reactive with NOVX protein or target molecules, but which do not interfere with binding of the NOVX protein to its target  
15 molecule, can be derivatized to the wells of the plate, and unbound target or NOVX protein trapped in the wells by antibody conjugation. Methods for detecting such complexes, in addition to those described above for the GST-immobilized complexes, include immunodetection of complexes using antibodies reactive with the NOVX protein or target molecule, as well as enzyme-linked assays that rely on detecting an enzymatic activity  
20 associated with the NOVX protein or target molecule.

In another embodiment, modulators of NOVX protein expression are identified in a method wherein a cell is contacted with a candidate compound and the expression of NOVX mRNA or protein in the cell is determined. The level of expression of NOVX mRNA or protein in the presence of the candidate compound is compared to the level of  
25 expression of NOVX mRNA or protein in the absence of the candidate compound. The candidate compound can then be identified as a modulator of NOVX mRNA or protein expression based upon this comparison. For example, when expression of NOVX mRNA or protein is greater (*i.e.*, statistically significantly greater) in the presence of the candidate compound than in its absence, the candidate compound is identified as a stimulator of  
30 NOVX mRNA or protein expression. Alternatively, when expression of NOVX mRNA or protein is less (statistically significantly less) in the presence of the candidate compound than in its absence, the candidate compound is identified as an inhibitor of NOVX mRNA

or protein expression. The level of NOVX mRNA or protein expression in the cells can be determined by methods described herein for detecting NOVX mRNA or protein.

In yet another aspect of the invention, the NOVX proteins can be used as "bait proteins" in a two-hybrid assay or three hybrid assay (*see, e.g.*, U.S. Patent No. 5,283,317; Zervos, *et al.*, 1993. *Cell* 72: 223-232; Madura, *et al.*, 1993. *J. Biol. Chem.* 268: 12046-12054; Bartel, *et al.*, 1993. *Biotechniques* 14: 920-924; Iwabuchi, *et al.*, 1993. *Oncogene* 8: 1693-1696; and Brent WO 94/10300), to identify other proteins that bind to or interact with NOVX ("NOVX-binding proteins" or "NOVX-bp") and modulate NOVX activity. Such NOVX-binding proteins are also involved in the propagation of signals by the NOVX proteins as, for example, upstream or downstream elements of the NOVX pathway.

The two-hybrid system is based on the modular nature of most transcription factors, which consist of separable DNA-binding and activation domains. Briefly, the assay utilizes two different DNA constructs. In one construct, the gene that codes for NOVX is fused to a gene encoding the DNA binding domain of a known transcription factor (*e.g.*, GAL-4). In the other construct, a DNA sequence, from a library of DNA sequences, that encodes an unidentified protein ("prey" or "sample") is fused to a gene that codes for the activation domain of the known transcription factor. If the "bait" and the "prey" proteins are able to interact, *in vivo*, forming a NOVX-dependent complex, the DNA-binding and activation domains of the transcription factor are brought into close proximity. This proximity allows transcription of a reporter gene (*e.g.*, LacZ) that is operably linked to a transcriptional regulatory site responsive to the transcription factor. Expression of the reporter gene can be detected and cell colonies containing the functional transcription factor can be isolated and used to obtain the cloned gene that encodes the protein which interacts with NOVX.

The invention further pertains to novel agents identified by the aforementioned screening assays and uses thereof for treatments as described herein.

### Detection Assays

Portions or fragments of the cDNA sequences identified herein (and the corresponding complete gene sequences) can be used in numerous ways as polynucleotide reagents. By way of example, and not of limitation, these sequences can be used to: (i) map their respective genes on a chromosome; and, thus, locate gene regions associated with genetic disease; (ii) identify an individual from a minute biological sample (tissue typing);

and (iii) aid in forensic identification of a biological sample. Some of these applications are described in the subsections, below.

### Chromosome Mapping

5        Once the sequence (or a portion of the sequence) of a gene has been isolated, this sequence can be used to map the location of the gene on a chromosome. This process is called chromosome mapping. Accordingly, portions or fragments of the NOVX sequences of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 226, or fragments or derivatives thereof, can be used to map the location of the NOVX genes, respectively, on a  
10 chromosome. The mapping of the NOVX sequences to chromosomes is an important first step in correlating these sequences with genes associated with disease.

      Briefly, NOVX genes can be mapped to chromosomes by preparing PCR primers (preferably 15-25 bp in length) from the NOVX sequences. Computer analysis of the NOVX sequences can be used to rapidly select primers that do not span more than one  
15 exon in the genomic DNA, thus complicating the amplification process. These primers can then be used for PCR screening of somatic cell hybrids containing individual human chromosomes. Only those hybrids containing the human gene corresponding to the NOVX sequences will yield an amplified fragment.

      Somatic cell hybrids are prepared by fusing somatic cells from different mammals  
20 (*e.g.*, human and mouse cells). As hybrids of human and mouse cells grow and divide, they gradually lose human chromosomes in random order, but retain the mouse chromosomes. By using media in which mouse cells cannot grow, because they lack a particular enzyme, but in which human cells can, the one human chromosome that contains the gene encoding the needed enzyme will be retained. By using various media, panels of hybrid cell lines  
25 can be established. Each cell line in a panel contains either a single human chromosome or a small number of human chromosomes, and a full set of mouse chromosomes, allowing easy mapping of individual genes to specific human chromosomes. *See, e.g.*, D'Eustachio, *et al.*, 1983. *Science* 220: 919-924. Somatic cell hybrids containing only fragments of human chromosomes can also be produced by using human chromosomes with  
30 translocations and deletions.

      PCR mapping of somatic cell hybrids is a rapid procedure for assigning a particular sequence to a particular chromosome. Three or more sequences can be assigned per day using a single thermal cycler. Using the NOVX sequences to design oligonucleotide

primers, sub-localization can be achieved with panels of fragments from specific chromosomes.

Fluorescence *in situ* hybridization (FISH) of a DNA sequence to a metaphase chromosomal spread can further be used to provide a precise chromosomal location in one step. Chromosome spreads can be made using cells whose division has been blocked in metaphase by a chemical like colcemid that disrupts the mitotic spindle. The chromosomes can be treated briefly with trypsin, and then stained with Giemsa. A pattern of light and dark bands develops on each chromosome, so that the chromosomes can be identified individually. The FISH technique can be used with a DNA sequence as short as 500 or 600 bases. However, clones larger than 1,000 bases have a higher likelihood of binding to a unique chromosomal location with sufficient signal intensity for simple detection. Preferably 1,000 bases, and more preferably 2,000 bases, will suffice to get good results at a reasonable amount of time. For a review of this technique, *see*, Verma, *et al.*, HUMAN CHROMOSOMES: A MANUAL OF BASIC TECHNIQUES (Pergamon Press, New York 1988).

Reagents for chromosome mapping can be used individually to mark a single chromosome or a single site on that chromosome, or panels of reagents can be used for marking multiple sites and/or multiple chromosomes. Reagents corresponding to noncoding regions of the genes actually are preferred for mapping purposes. Coding sequences are more likely to be conserved within gene families, thus increasing the chance of cross hybridizations during chromosomal mapping.

Once a sequence has been mapped to a precise chromosomal location, the physical position of the sequence on the chromosome can be correlated with genetic map data. Such data are found, *e.g.*, in McKusick, MENDELIAN INHERITANCE IN MAN, available on-line through Johns Hopkins University Welch Medical Library). The relationship between genes and disease, mapped to the same chromosomal region, can then be identified through linkage analysis (co-inheritance of physically adjacent genes), described in, *e.g.*, Egeland, *et al.*, 1987. *Nature*, 325: 783-787.

Moreover, differences in the DNA sequences between individuals affected and unaffected with a disease associated with the NOVX gene, can be determined. If a mutation is observed in some or all of the affected individuals but not in any unaffected individuals, then the mutation is likely to be the causative agent of the particular disease. Comparison of affected and unaffected individuals generally involves first looking for structural alterations in the chromosomes, such as deletions or translocations that are

visible from chromosome spreads or detectable using PCR based on that DNA sequence. Ultimately, complete sequencing of genes from several individuals can be performed to confirm the presence of a mutation and to distinguish mutations from polymorphisms.

## 5 Tissue Typing

The NOVX sequences of the invention can also be used to identify individuals from minute biological samples. In this technique, an individual's genomic DNA is digested with one or more restriction enzymes, and probed on a Southern blot to yield unique bands for identification. The sequences of the invention are useful as additional DNA markers for  
10 RFLP ("restriction fragment length polymorphisms," described in U.S. Patent No. 5,272,057).

Furthermore, the sequences of the invention can be used to provide an alternative technique that determines the actual base-by-base DNA sequence of selected portions of an individual's genome. Thus, the NOVX sequences described herein can be used to prepare  
15 two PCR primers from the 5'- and 3'-termini of the sequences. These primers can then be used to amplify an individual's DNA and subsequently sequence it.

Panels of corresponding DNA sequences from individuals, prepared in this manner, can provide unique individual identifications, as each individual will have a unique set of such DNA sequences due to allelic differences. The sequences of the invention can be used  
20 to obtain such identification sequences from individuals and from tissue. The NOVX sequences of the invention uniquely represent portions of the human genome. Allelic variation occurs to some degree in the coding regions of these sequences, and to a greater degree in the noncoding regions. It is estimated that allelic variation between individual humans occurs with a frequency of about once per each 500 bases. Much of the allelic  
25 variation is due to single nucleotide polymorphisms (SNPs), which include restriction fragment length polymorphisms (RFLPs).

Each of the sequences described herein can, to some degree, be used as a standard against which DNA from an individual can be compared for identification purposes. Because greater numbers of polymorphisms occur in the noncoding regions, fewer  
30 sequences are necessary to differentiate individuals. The noncoding sequences can comfortably provide positive individual identification with a panel of perhaps 10 to 1,000 primers that each yield a noncoding amplified sequence of 100 bases. If coding sequences, such as those of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 226, are used, a

more appropriate number of primers for positive individual identification would be 500-2,000.

### Predictive Medicine

5           The invention also pertains to the field of predictive medicine in which diagnostic assays, prognostic assays, pharmacogenomics, and monitoring clinical trials are used for prognostic (predictive) purposes to thereby treat an individual prophylactically. Accordingly, one aspect of the invention relates to diagnostic assays for determining NOVX protein and/or nucleic acid expression as well as NOVX activity, in the context of a  
10   biological sample (*e.g.*, blood, serum, cells, tissue) to thereby determine whether an individual is afflicted with a disease or disorder, or is at risk of developing a disorder, associated with aberrant NOVX expression or activity. The disorders include metabolic disorders, diabetes, obesity, infectious disease, anorexia, cancer-associated cachexia, cancer, neurodegenerative disorders, Alzheimer's Disease, Parkinson's Disorder, immune  
15   disorders, and hematopoietic disorders, and the various dyslipidemias, metabolic disturbances associated with obesity, the metabolic syndrome X and wasting disorders associated with chronic diseases and various cancers. The invention also provides for prognostic (or predictive) assays for determining whether an individual is at risk of developing a disorder associated with NOVX protein, nucleic acid expression or activity.  
20   For example, mutations in a NOVX gene can be assayed in a biological sample. Such assays can be used for prognostic or predictive purpose to thereby prophylactically treat an individual prior to the onset of a disorder characterized by or associated with NOVX protein, nucleic acid expression, or biological activity.

          Another aspect of the invention provides methods for determining NOVX protein,  
25   nucleic acid expression or activity in an individual to thereby select appropriate therapeutic or prophylactic agents for that individual (referred to herein as "pharmacogenomics"). Pharmacogenomics allows for the selection of agents (*e.g.*, drugs) for therapeutic or prophylactic treatment of an individual based on the genotype of the individual (*e.g.*, the genotype of the individual examined to determine the ability of the individual to respond to  
30   a particular agent.)

          Yet another aspect of the invention pertains to monitoring the influence of agents (*e.g.*, drugs, compounds) on the expression or activity of NOVX in clinical trials.

          These and other agents are described in further detail in the following sections.

### Diagnostic Assays

An exemplary method for detecting the presence or absence of NOVX in a biological sample involves obtaining a biological sample from a test subject and contacting the biological sample with a compound or an agent capable of detecting NOVX protein or nucleic acid (*e.g.*, mRNA, genomic DNA) that encodes NOVX protein such that the presence of NOVX is detected in the biological sample. An agent for detecting NOVX mRNA or genomic DNA is a labeled nucleic acid probe capable of hybridizing to NOVX mRNA or genomic DNA. The nucleic acid probe can be, for example, a full-length NOVX nucleic acid, such as the nucleic acid of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 226, or a portion thereof, such as an oligonucleotide of at least 15, 30, 50, 100, 250 or 500 nucleotides in length and sufficient to specifically hybridize under stringent conditions to NOVX mRNA or genomic DNA. Other suitable probes for use in the diagnostic assays of the invention are described herein.

An agent for detecting NOVX protein is an antibody capable of binding to NOVX protein, preferably an antibody with a detectable label. Antibodies can be polyclonal, or more preferably, monoclonal. An intact antibody, or a fragment thereof (*e.g.*, Fab or F(ab')<sub>2</sub>) can be used. The term "labeled", with regard to the probe or antibody, is intended to encompass direct labeling of the probe or antibody by coupling (*i.e.*, physically linking) a detectable substance to the probe or antibody, as well as indirect labeling of the probe or antibody by reactivity with another reagent that is directly labeled. Examples of indirect labeling include detection of a primary antibody using a fluorescently-labeled secondary antibody and end-labeling of a DNA probe with biotin such that it can be detected with fluorescently-labeled streptavidin. The term "biological sample" is intended to include tissues, cells and biological fluids isolated from a subject, as well as tissues, cells and fluids present within a subject. That is, the detection method of the invention can be used to detect NOVX mRNA, protein, or genomic DNA in a biological sample *in vitro* as well as *in vivo*. For example, *in vitro* techniques for detection of NOVX mRNA include Northern hybridizations and *in situ* hybridizations. *In vitro* techniques for detection of NOVX protein include enzyme linked immunosorbent assays (ELISAs), Western blots, immunoprecipitations, and immunofluorescence. *In vitro* techniques for detection of NOVX genomic DNA include Southern hybridizations. Furthermore, *in vivo* techniques for detection of NOVX protein include introducing into a subject a labeled anti-NOVX

antibody. For example, the antibody can be labeled with a radioactive marker whose presence and location in a subject can be detected by standard imaging techniques.

In one embodiment, the biological sample contains protein molecules from the test subject. Alternatively, the biological sample can contain mRNA molecules from the test  
5 subject or genomic DNA molecules from the test subject. A preferred biological sample is a peripheral blood leukocyte sample isolated by conventional means from a subject.

In another embodiment, the methods further involve obtaining a control biological sample from a control subject, contacting the control sample with a compound or agent capable of detecting NOVX protein, mRNA, or genomic DNA, such that the presence of  
10 NOVX protein, mRNA or genomic DNA is detected in the biological sample, and comparing the presence of NOVX protein, mRNA or genomic DNA in the control sample with the presence of NOVX protein, mRNA or genomic DNA in the test sample.

The invention also encompasses kits for detecting the presence of NOVX in a biological sample. For example, the kit can comprise: a labeled compound or agent  
15 capable of detecting NOVX protein or mRNA in a biological sample; means for determining the amount of NOVX in the sample; and means for comparing the amount of NOVX in the sample with a standard. The compound or agent can be packaged in a suitable container. The kit can further comprise instructions for using the kit to detect NOVX protein or nucleic acid.

20

### **Prognostic Assays**

The diagnostic methods described herein can furthermore be utilized to identify subjects having or at risk of developing a disease or disorder associated with aberrant NOVX expression or activity. For example, the assays described herein, such as the  
25 preceding diagnostic assays or the following assays, can be utilized to identify a subject having or at risk of developing a disorder associated with NOVX protein, nucleic acid expression or activity. Alternatively, the prognostic assays can be utilized to identify a subject having or at risk for developing a disease or disorder. Thus, the invention provides a method for identifying a disease or disorder associated with aberrant NOVX expression  
30 or activity in which a test sample is obtained from a subject and NOVX protein or nucleic acid (*e.g.*, mRNA, genomic DNA) is detected, wherein the presence of NOVX protein or nucleic acid is diagnostic for a subject having or at risk of developing a disease or disorder associated with aberrant NOVX expression or activity. As used herein, a "test sample"

refers to a biological sample obtained from a subject of interest. For example, a test sample can be a biological fluid (*e.g.*, serum), cell sample, or tissue.

Furthermore, the prognostic assays described herein can be used to determine whether a subject can be administered an agent (*e.g.*, an agonist, antagonist, peptidomimetic, protein, peptide, nucleic acid, small molecule, or other drug candidate) to treat a disease or disorder associated with aberrant NOVX expression or activity. For example, such methods can be used to determine whether a subject can be effectively treated with an agent for a disorder. Thus, the invention provides methods for determining whether a subject can be effectively treated with an agent for a disorder associated with aberrant NOVX expression or activity in which a test sample is obtained and NOVX protein or nucleic acid is detected (*e.g.*, wherein the presence of NOVX protein or nucleic acid is diagnostic for a subject that can be administered the agent to treat a disorder associated with aberrant NOVX expression or activity).

The methods of the invention can also be used to detect genetic lesions in a NOVX gene, thereby determining if a subject with the lesioned gene is at risk for a disorder characterized by aberrant cell proliferation and/or differentiation. In various embodiments, the methods include detecting, in a sample of cells from the subject, the presence or absence of a genetic lesion characterized by at least one of an alteration affecting the integrity of a gene encoding a NOVX-protein, or the misexpression of the NOVX gene. For example, such genetic lesions can be detected by ascertaining the existence of at least one of: (i) a deletion of one or more nucleotides from a NOVX gene; (ii) an addition of one or more nucleotides to a NOVX gene; (iii) a substitution of one or more nucleotides of a NOVX gene, (iv) a chromosomal rearrangement of a NOVX gene; (v) an alteration in the level of a messenger RNA transcript of a NOVX gene, (vi) aberrant modification of a NOVX gene, such as of the methylation pattern of the genomic DNA, (vii) the presence of a non-wild-type splicing pattern of a messenger RNA transcript of a NOVX gene, (viii) a non-wild-type level of a NOVX protein, (ix) allelic loss of a NOVX gene, and (x) inappropriate post-translational modification of a NOVX protein. As described herein, there are a large number of assay techniques known in the art which can be used for detecting lesions in a NOVX gene. A preferred biological sample is a peripheral blood leukocyte sample isolated by conventional means from a subject. However, any biological sample containing nucleated cells may be used, including, for example, buccal mucosal cells.

In certain embodiments, detection of the lesion involves the use of a probe/primer in a polymerase chain reaction (PCR) (*see, e.g.*, U.S. Patent Nos. 4,683,195 and 4,683,202), such as anchor PCR or RACE PCR, or, alternatively, in a ligation chain reaction (LCR) (*see, e.g.*, Landegran, *et al.*, 1988. *Science* 241: 1077-1080; and Nakazawa, *et al.*, 1994. *Proc. Natl. Acad. Sci. USA* 91: 360-364), the latter of which can be particularly useful for detecting point mutations in the NOVX-gene (*see*, Abravaya, *et al.*, 1995. *Nucl. Acids Res.* 23: 675-682). This method can include the steps of collecting a sample of cells from a patient, isolating nucleic acid (*e.g.*, genomic, mRNA or both) from the cells of the sample, contacting the nucleic acid sample with one or more primers that specifically hybridize to a NOVX gene under conditions such that hybridization and amplification of the NOVX gene (if present) occurs, and detecting the presence or absence of an amplification product, or detecting the size of the amplification product and comparing the length to a control sample. It is anticipated that PCR and/or LCR may be desirable to use as a preliminary amplification step in conjunction with any of the techniques used for detecting mutations described herein.

Alternative amplification methods include: self sustained sequence replication (*see*, Guatelli, *et al.*, 1990. *Proc. Natl. Acad. Sci. USA* 87: 1874-1878), transcriptional amplification system (*see*, Kwoh, *et al.*, 1989. *Proc. Natl. Acad. Sci. USA* 86: 1173-1177); Q $\beta$  Replicase (*see*, Lizardi, *et al.*, 1988. *BioTechnology* 6: 1197), or any other nucleic acid amplification method, followed by the detection of the amplified molecules using techniques well known to those of skill in the art. These detection schemes are especially useful for the detection of nucleic acid molecules if such molecules are present in very low numbers.

In an alternative embodiment, mutations in a NOVX gene from a sample cell can be identified by alterations in restriction enzyme cleavage patterns. For example, sample and control DNA is isolated, amplified (optionally), digested with one or more restriction endonucleases, and fragment length sizes are determined by gel electrophoresis and compared. Differences in fragment length sizes between sample and control DNA indicates mutations in the sample DNA. Moreover, the use of sequence specific ribozymes (*see, e.g.*, U.S. Patent No. 5,493,531) can be used to score for the presence of specific mutations by development or loss of a ribozyme cleavage site.

In other embodiments, genetic mutations in NOVX can be identified by hybridizing a sample and control nucleic acids, *e.g.*, DNA or RNA, to high-density arrays containing

hundreds or thousands of oligonucleotides probes. *See, e.g., Cronin, et al., 1996. Human Mutation* 7: 244-255; Kozal, *et al., 1996. Nat. Med.* 2: 753-759. For example, genetic mutations in NOVX can be identified in two dimensional arrays containing light-generated DNA probes as described in Cronin, *et al., supra*. Briefly, a first hybridization array of probes can be used to scan through long stretches of DNA in a sample and control to identify base changes between the sequences by making linear arrays of sequential overlapping probes. This step allows the identification of point mutations. This is followed by a second hybridization array that allows the characterization of specific mutations by using smaller, specialized probe arrays complementary to all variants or mutations detected. Each mutation array is composed of parallel probe sets, one complementary to the wild-type gene and the other complementary to the mutant gene.

In yet another embodiment, any of a variety of sequencing reactions known in the art can be used to directly sequence the NOVX gene and detect mutations by comparing the sequence of the sample NOVX with the corresponding wild-type (control) sequence. Examples of sequencing reactions include those based on techniques developed by Maxim and Gilbert, 1977. *Proc. Natl. Acad. Sci. USA* 74: 560 or Sanger, 1977. *Proc. Natl. Acad. Sci. USA* 74: 5463. It is also contemplated that any of a variety of automated sequencing procedures can be utilized when performing the diagnostic assays (*see, e.g., Naeve, et al., 1995. Biotechniques* 19: 448), including sequencing by mass spectrometry (*see, e.g., PCT International Publication No. WO 94/16101; Cohen, et al., 1996. Adv. Chromatography* 36: 127-162; and Griffin, *et al., 1993. Appl. Biochem. Biotechnol.* 38: 147-159).

Other methods for detecting mutations in the NOVX gene include methods in which protection from cleavage agents is used to detect mismatched bases in RNA/RNA or RNA/DNA heteroduplexes. *See, e.g., Myers, et al., 1985. Science* 230: 1242. In general, the art technique of "mismatch cleavage" starts by providing heteroduplexes of formed by hybridizing (labeled) RNA or DNA containing the wild-type NOVX sequence with potentially mutant RNA or DNA obtained from a tissue sample. The double-stranded duplexes are treated with an agent that cleaves single-stranded regions of the duplex such as which will exist due to basepair mismatches between the control and sample strands. For instance, RNA/DNA duplexes can be treated with RNase and DNA/DNA hybrids treated with S<sub>1</sub> nuclease to enzymatically digesting the mismatched regions. In other embodiments, either DNA/DNA or RNA/DNA duplexes can be treated with hydroxylamine or osmium tetroxide and with piperidine in order to digest mismatched

regions. After digestion of the mismatched regions, the resulting material is then separated by size on denaturing polyacrylamide gels to determine the site of mutation. *See, e.g.,* Cotton, *et al.*, 1988. *Proc. Natl. Acad. Sci. USA* 85: 4397; Saleeba, *et al.*, 1992. *Methods Enzymol.* 217: 286-295. In an embodiment, the control DNA or RNA can be labeled for  
5 detection.

In still another embodiment, the mismatch cleavage reaction employs one or more proteins that recognize mismatched base pairs in double-stranded DNA (so called "DNA mismatch repair" enzymes) in defined systems for detecting and mapping point mutations in NOVX cDNAs obtained from samples of cells. For example, the mutY enzyme of *E.*  
10 *coli* cleaves A at G/A mismatches and the thymidine DNA glycosylase from HeLa cells cleaves T at G/T mismatches. *See, e.g.,* Hsu, *et al.*, 1994. *Carcinogenesis* 15: 1657-1662. According to an exemplary embodiment, a probe based on a NOVX sequence, *e.g.,* a wild-type NOVX sequence, is hybridized to a cDNA or other DNA product from a test cell(s). The duplex is treated with a DNA mismatch repair enzyme, and the cleavage  
15 products, if any, can be detected from electrophoresis protocols or the like. *See, e.g.,* U.S. Patent No. 5,459,039.

In other embodiments, alterations in electrophoretic mobility will be used to identify mutations in NOVX genes. For example, single strand conformation polymorphism (SSCP) may be used to detect differences in electrophoretic mobility  
20 between mutant and wild type nucleic acids. *See, e.g.,* Orita, *et al.*, 1989. *Proc. Natl. Acad. Sci. USA*: 86: 2766; Cotton, 1993. *Mutat. Res.* 285: 125-144; Hayashi, 1992. *Genet. Anal. Tech. Appl.* 9: 73-79. Single-stranded DNA fragments of sample and control NOVX nucleic acids will be denatured and allowed to renature. The secondary structure of single-stranded nucleic acids varies according to sequence, the resulting alteration in  
25 electrophoretic mobility enables the detection of even a single base change. The DNA fragments may be labeled or detected with labeled probes. The sensitivity of the assay may be enhanced by using RNA (rather than DNA), in which the secondary structure is more sensitive to a change in sequence. In one embodiment, the subject method utilizes heteroduplex analysis to separate double stranded heteroduplex molecules on the basis of  
30 changes in electrophoretic mobility. *See, e.g.,* Keen, *et al.*, 1991. *Trends Genet.* 7: 5.

In yet another embodiment, the movement of mutant or wild-type fragments in polyacrylamide gels containing a gradient of denaturant is assayed using denaturing gradient gel electrophoresis (DGGE). *See, e.g.,* Myers, *et al.*, 1985. *Nature* 313: 495.

When DGGE is used as the method of analysis, DNA will be modified to insure that it does not completely denature, for example by adding a GC clamp of approximately 40 bp of high-melting GC-rich DNA by PCR. In a further embodiment, a temperature gradient is used in place of a denaturing gradient to identify differences in the mobility of control and sample DNA. See, e.g., Rosenbaum and Reissner, 1987. *Biophys. Chem.* 265: 12753.

Examples of other techniques for detecting point mutations include, but are not limited to, selective oligonucleotide hybridization, selective amplification, or selective primer extension. For example, oligonucleotide primers may be prepared in which the known mutation is placed centrally and then hybridized to target DNA under conditions that permit hybridization only if a perfect match is found. See, e.g., Saiki, *et al.*, 1986. *Nature* 324: 163; Saiki, *et al.*, 1989. *Proc. Natl. Acad. Sci. USA* 86: 6230. Such allele specific oligonucleotides are hybridized to PCR amplified target DNA or a number of different mutations when the oligonucleotides are attached to the hybridizing membrane and hybridized with labeled target DNA.

Alternatively, allele specific amplification technology that depends on selective PCR amplification may be used in conjunction with the instant invention. Oligonucleotides used as primers for specific amplification may carry the mutation of interest in the center of the molecule (so that amplification depends on differential hybridization; see, e.g., Gibbs, *et al.*, 1989. *Nucl. Acids Res.* 17: 2437-2448) or at the extreme 3'-terminus of one primer where, under appropriate conditions, mismatch can prevent, or reduce polymerase extension (see, e.g., Prossner, 1993. *Tibtech.* 11: 238). In addition it may be desirable to introduce a novel restriction site in the region of the mutation to create cleavage-based detection. See, e.g., Gasparini, *et al.*, 1992. *Mol. Cell Probes* 6: 1. It is anticipated that in certain embodiments amplification may also be performed using *Taq* ligase for amplification. See, e.g., Barany, 1991. *Proc. Natl. Acad. Sci. USA* 88: 189. In such cases, ligation will occur only if there is a perfect match at the 3'-terminus of the 5' sequence, making it possible to detect the presence of a known mutation at a specific site by looking for the presence or absence of amplification.

The methods described herein may be performed, for example, by utilizing pre-packaged diagnostic kits comprising at least one probe nucleic acid or antibody reagent described herein, which may be conveniently used, e.g., in clinical settings to diagnose patients exhibiting symptoms or family history of a disease or illness involving a NOVX gene.

Furthermore, any cell type or tissue, preferably peripheral blood leukocytes, in which NOVX is expressed may be utilized in the prognostic assays described herein. However, any biological sample containing nucleated cells may be used, including, for example, buccal mucosal cells.

5

### Pharmacogenomics

Agents, or modulators that have a stimulatory or inhibitory effect on NOVX activity (*e.g.*, NOVX gene expression), as identified by a screening assay described herein can be administered to individuals to treat (prophylactically or therapeutically) disorders. The disorders include but are not limited to, *e.g.*, those diseases, disorders and conditions listed above, and more particularly include those diseases, disorders, or conditions associated with homologs of a NOVX protein, such as those summarized in Table A.

In conjunction with such treatment, the pharmacogenomics (*i.e.*, the study of the relationship between an individual's genotype and that individual's response to a foreign compound or drug) of the individual may be considered. Differences in metabolism of therapeutics can lead to severe toxicity or therapeutic failure by altering the relation between dose and blood concentration of the pharmacologically active drug. Thus, the pharmacogenomics of the individual permits the selection of effective agents (*e.g.*, drugs) for prophylactic or therapeutic treatments based on a consideration of the individual's genotype. Such pharmacogenomics can further be used to determine appropriate dosages and therapeutic regimens. Accordingly, the activity of NOVX protein, expression of NOVX nucleic acid, or mutation content of NOVX genes in an individual can be determined to thereby select appropriate agent(s) for therapeutic or prophylactic treatment of the individual.

Pharmacogenomics deals with clinically significant hereditary variations in the response to drugs due to altered drug disposition and abnormal action in affected persons. See *e.g.*, Eichelbaum, 1996. *Clin. Exp. Pharmacol. Physiol.*, 23: 983-985; Linder, 1997. *Clin. Chem.*, 43: 254-266. In general, two types of pharmacogenetic conditions can be differentiated. Genetic conditions transmitted as a single factor altering the way drugs act on the body (altered drug action) or genetic conditions transmitted as single factors altering the way the body acts on drugs (altered drug metabolism). These pharmacogenetic conditions can occur either as rare defects or as polymorphisms. For example, glucose-6-phosphate dehydrogenase (G6PD) deficiency is a common inherited

enzymopathy in which the main clinical complication is hemolysis after ingestion of oxidant drugs (anti-malarials, sulfonamides, analgesics, nitrofurans) and consumption of fava beans.

As an illustrative embodiment, the activity of drug metabolizing enzymes is a major determinant of both the intensity and duration of drug action. The discovery of genetic polymorphisms of drug metabolizing enzymes (*e.g.*, N-acetyltransferase 2 (NAT 2) and cytochrome pregnancy zone protein precursor enzymes CYP2D6 and CYP2C19) has provided an explanation as to why some patients do not obtain the expected drug effects or show exaggerated drug response and serious toxicity after taking the standard and safe dose of a drug. These polymorphisms are expressed in two phenotypes in the population, the extensive metabolizer (EM) and poor metabolizer (PM). The prevalence of PM is different among different populations. For example, the gene coding for CYP2D6 is highly polymorphic and several mutations have been identified in PM, which all lead to the absence of functional CYP2D6. Poor metabolizers of CYP2D6 and CYP2C19 quite frequently experience exaggerated drug response and side effects when they receive standard doses. If a metabolite is the active therapeutic moiety, PM show no therapeutic response, as demonstrated for the analgesic effect of codeine mediated by its CYP2D6-formed metabolite morphine. At the other extreme are the so called ultra-rapid metabolizers who do not respond to standard doses. Recently, the molecular basis of ultra-rapid metabolism has been identified to be due to CYP2D6 gene amplification.

Thus, the activity of NOVX protein, expression of NOVX nucleic acid, or mutation content of NOVX genes in an individual can be determined to thereby select appropriate agent(s) for therapeutic or prophylactic treatment of the individual. In addition, pharmacogenetic studies can be used to apply genotyping of polymorphic alleles encoding drug-metabolizing enzymes to the identification of an individual's drug responsiveness phenotype. This knowledge, when applied to dosing or drug selection, can avoid adverse reactions or therapeutic failure and thus enhance therapeutic or prophylactic efficiency when treating a subject with a NOVX modulator, such as a modulator identified by one of the exemplary screening assays described herein.

30

#### Monitoring of Effects During Clinical Trials

Monitoring the influence of agents (*e.g.*, drugs, compounds) on the expression or activity of NOVX (*e.g.*, the ability to modulate aberrant cell proliferation and/or

differentiation) can be applied not only in basic drug screening, but also in clinical trials. For example, the effectiveness of an agent determined by a screening assay as described herein to increase NOVX gene expression, protein levels, or upregulate NOVX activity, can be monitored in clinical trials of subjects exhibiting decreased NOVX gene expression, protein levels, or downregulated NOVX activity. Alternatively, the effectiveness of an agent determined by a screening assay to decrease NOVX gene expression, protein levels, or downregulate NOVX activity, can be monitored in clinical trials of subjects exhibiting increased NOVX gene expression, protein levels, or upregulated NOVX activity. In such clinical trials, the expression or activity of NOVX and, preferably, other genes that have been implicated in, for example, a cellular proliferation or immune disorder can be used as a "read out" or markers of the immune responsiveness of a particular cell.

By way of example, and not of limitation, genes, including NOVX, that are modulated in cells by treatment with an agent (*e.g.*, compound, drug or small molecule) that modulates NOVX activity (*e.g.*, identified in a screening assay as described herein) can be identified. Thus, to study the effect of agents on cellular proliferation disorders, for example, in a clinical trial, cells can be isolated and RNA prepared and analyzed for the levels of expression of NOVX and other genes implicated in the disorder. The levels of gene expression (*i.e.*, a gene expression pattern) can be quantified by Northern blot analysis or RT-PCR, as described herein, or alternatively by measuring the amount of protein produced, by one of the methods as described herein, or by measuring the levels of activity of NOVX or other genes. In this manner, the gene expression pattern can serve as a marker, indicative of the physiological response of the cells to the agent. Accordingly, this response state may be determined before, and at various points during, treatment of the individual with the agent.

In one embodiment, the invention provides a method for monitoring the effectiveness of treatment of a subject with an agent (*e.g.*, an agonist, antagonist, protein, peptide, peptidomimetic, nucleic acid, small molecule, or other drug candidate identified by the screening assays described herein) comprising the steps of (i) obtaining a pre-administration sample from a subject prior to administration of the agent; (ii) detecting the level of expression of a NOVX protein, mRNA, or genomic DNA in the preadministration sample; (iii) obtaining one or more post-administration samples from the subject; (iv) detecting the level of expression or activity of the NOVX protein, mRNA, or genomic DNA in the post-administration samples; (v) comparing the level of expression or

activity of the NOVX protein, mRNA, or genomic DNA in the pre-administration sample with the NOVX protein, mRNA, or genomic DNA in the post administration sample or samples; and (vi) altering the administration of the agent to the subject accordingly. For example, increased administration of the agent may be desirable to increase the expression or activity of NOVX to higher levels than detected, *i.e.*, to increase the effectiveness of the agent. Alternatively, decreased administration of the agent may be desirable to decrease expression or activity of NOVX to lower levels than detected, *i.e.*, to decrease the effectiveness of the agent.

## 10 Methods of Treatment

The invention provides for both prophylactic and therapeutic methods of treating a subject at risk of (or susceptible to) a disorder or having a disorder associated with aberrant NOVX expression or activity. The disorders include but are not limited to, *e.g.*, those diseases, disorders and conditions listed above, and more particularly include those diseases, disorders, or conditions associated with homologs of a NOVX protein, such as those summarized in Table A.

These methods of treatment will be discussed more fully, below.

### Diseases and Disorders

Diseases and disorders that are characterized by increased (relative to a subject not suffering from the disease or disorder) levels or biological activity may be treated with Therapeutics that antagonize (*i.e.*, reduce or inhibit) activity. Therapeutics that antagonize activity may be administered in a therapeutic or prophylactic manner. Therapeutics that may be utilized include, but are not limited to: (i) an aforementioned peptide, or analogs, derivatives, fragments or homologs thereof; (ii) antibodies to an aforementioned peptide; (iii) nucleic acids encoding an aforementioned peptide; (iv) administration of antisense nucleic acid and nucleic acids that are "dysfunctional" (*i.e.*, due to a heterologous insertion within the coding sequences of coding sequences to an aforementioned peptide) that are utilized to "knockout" endogenous function of an aforementioned peptide by homologous recombination (*see, e.g.*, Capecchi, 1989. *Science* 244: 1288-1292); or (v) modulators (*i.e.*, inhibitors, agonists and antagonists, including additional peptide mimetic of the invention or antibodies specific to a peptide of the invention) that alter the interaction between an aforementioned peptide and its binding partner.

Diseases and disorders that are characterized by decreased (relative to a subject not suffering from the disease or disorder) levels or biological activity may be treated with Therapeutics that increase (*i.e.*, are agonists to) activity. Therapeutics that upregulate activity may be administered in a therapeutic or prophylactic manner. Therapeutics that  
5 may be utilized include, but are not limited to, an aforementioned peptide, or analogs, derivatives, fragments or homologs thereof; or an agonist that increases bioavailability.

Increased or decreased levels can be readily detected by quantifying peptide and/or RNA, by obtaining a patient tissue sample (*e.g.*, from biopsy tissue) and assaying it *in vitro* for RNA or peptide levels, structure and/or activity of the expressed peptides (or mRNAs  
10 of an aforementioned peptide). Methods that are well-known within the art include, but are not limited to, immunoassays (*e.g.*, by Western blot analysis, immunoprecipitation followed by sodium dodecyl sulfate (SDS) polyacrylamide gel electrophoresis, immunocytochemistry, *etc.*) and/or hybridization assays to detect expression of mRNAs (*e.g.*, Northern assays, dot blots, *in situ* hybridization, and the like).

15

#### **Prophylactic Methods**

In one aspect, the invention provides a method for preventing, in a subject, a disease or condition associated with an aberrant NOVX expression or activity, by administering to the subject an agent that modulates NOVX expression or at least one NOVX activity.  
20 Subjects at risk for a disease that is caused or contributed to by aberrant NOVX expression or activity can be identified by, for example, any or a combination of diagnostic or prognostic assays as described herein. Administration of a prophylactic agent can occur prior to the manifestation of symptoms characteristic of the NOVX aberrancy, such that a disease or disorder is prevented or, alternatively, delayed in its progression. Depending  
25 upon the type of NOVX aberrancy, for example, a NOVX agonist or NOVX antagonist agent can be used for treating the subject. The appropriate agent can be determined based on screening assays described herein. The prophylactic methods of the invention are further discussed in the following subsections.

#### **Therapeutic Methods**

Another aspect of the invention pertains to methods of modulating NOVX expression or activity for therapeutic purposes. The modulatory method of the invention involves contacting a cell with an agent that modulates one or more of the activities of

NOVX protein activity associated with the cell. An agent that modulates NOVX protein activity can be an agent as described herein, such as a nucleic acid or a protein, a naturally-occurring cognate ligand of a NOVX protein, a peptide, a NOVX peptidomimetic, or other small molecule. In one embodiment, the agent stimulates one or more NOVX protein activity. Examples of such stimulatory agents include active NOVX protein and a nucleic acid molecule encoding NOVX that has been introduced into the cell. In another embodiment, the agent inhibits one or more NOVX protein activity. Examples of such inhibitory agents include antisense NOVX nucleic acid molecules and anti-NOVX antibodies. These modulatory methods can be performed *in vitro* (e.g., by culturing the cell with the agent) or, alternatively, *in vivo* (e.g., by administering the agent to a subject). As such, the invention provides methods of treating an individual afflicted with a disease or disorder characterized by aberrant expression or activity of a NOVX protein or nucleic acid molecule. In one embodiment, the method involves administering an agent (e.g., an agent identified by a screening assay described herein), or combination of agents that modulates (e.g., up-regulates or down-regulates) NOVX expression or activity. In another embodiment, the method involves administering a NOVX protein or nucleic acid molecule as therapy to compensate for reduced or aberrant NOVX expression or activity.

Stimulation of NOVX activity is desirable *in situations* in which NOVX is abnormally downregulated and/or in which increased NOVX activity is likely to have a beneficial effect. One example of such a situation is where a subject has a disorder characterized by aberrant cell proliferation and/or differentiation (e.g., cancer or immune associated disorders). Another example of such a situation is where the subject has a gestational disease (e.g., preclampsia).

## **Determination of the Biological Effect of the Therapeutic**

In various embodiments of the invention, suitable *in vitro* or *in vivo* assays are performed to determine the effect of a specific Therapeutic and whether its administration is indicated for treatment of the affected tissue.

In various specific embodiments, *in vitro* assays may be performed with representative cells of the type(s) involved in the patient's disorder, to determine if a given Therapeutic exerts the desired effect upon the cell type(s). Compounds for use in therapy may be tested in suitable animal model systems including, but not limited to rats, mice, chicken, cows, monkeys, rabbits, and the like, prior to testing in human subjects. Similarly,

for *in vivo* testing, any of the animal model system known in the art may be used prior to administration to human subjects.

#### **Prophylactic and Therapeutic Uses of the Compositions of the Invention**

5           The NOVX nucleic acids and proteins of the invention are useful in potential prophylactic and therapeutic applications implicated in a variety of disorders. The disorders include but are not limited to, *e.g.*, those diseases, disorders and conditions listed above, and more particularly include those diseases, disorders, or conditions associated with homologs of a NOVX protein, such as those summarized in Table A.

10           As an example, a cDNA encoding the NOVX protein of the invention may be useful in gene therapy, and the protein may be useful when administered to a subject in need thereof. By way of non-limiting example, the compositions of the invention will have efficacy for treatment of patients suffering from diseases, disorders, conditions and the like, including but not limited to those listed herein.

15           Both the novel nucleic acid encoding the NOVX protein, and the NOVX protein of the invention, or fragments thereof, may also be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed. A further use could be as an anti-bacterial molecule (*i.e.*, some peptides have been found to possess anti-bacterial properties). These materials are further useful in the generation of antibodies,  
20           which immunospecifically-bind to the novel substances of the invention for use in therapeutic or diagnostic methods.

          The invention will be further described in the following examples, which do not limit the scope of the invention described in the claims.

## EXAMPLES

## Example A. Polynucleotide and Polypeptide Sequences, and Homology Data

The NOV1 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 1A.

5

Table 1A. NOV1 Sequence Analysis			
	SEQ ID NO: 1	2763 bp	
NOV1a, CG101683-01 DNA Sequence	GGATCCCAGTGGCCCGCGTGCCTCGGCTCCACAGGCCTGCAGCCAGCATCGCACCGA		
	ACCTTCGGGGGGCCGCGGCTGGAGCGCTCGGCCGGCGTGGGAGCCGAAGGCCGAGAT		
	GCAATCTTCTTACCGCGAAGAAGCCAGGGGAATAGGTAGCCACATCTTGTGTGCAGAT		
	AAGAAAGGAAGCTAACGCAGTATCTGCAAAGCCAGGAGTCTGACTCAGTACTTTTCTC		
	ACTCATGCATACAAGCAGCTAAAAATGACACAGCTTATTTACCATGCCCTGACACTG		
	CACTGAGCACTTTATGAGCTTGAACCTCTGTTAATCTCACGACCACCTCATGAGACTCT		
	CCAGAAAGAGCAACAGTAATGGAGTACATGAGCACTGGAAGTGACAATAAAGAAGAGA		
	TTGATTTATTAATTAAACATTTAAATGTGTCTGATGTAATAGACATTATGAAAAATCT		
	TTATGCAAGTGAAGAGCCAGCAGTTTATGAACCCAGTCTAATGACCATGTGTCAAGAC		
	AGTAATCAAAACGATGAGCGTTCTAAGTCTCTGCTGCTTAGTGGCCAAGAGGTACCAT		
	GGTTGTCATCAGTCAGATATGGAAGTGTGGAGGATTTGCTTGCTTTTGCAAACCATAT		
	ATCCAACACTGCAAAGCATTTTATGGACAACGACCACAGGAATCTGGAATTTTATTA		
	AACATGGTCATCACTCCCCAAAATGGACGTTACCAAATAGATTCGATGTTCTCCTGA		
	TCCCTTGGAAAGCTGACTTACAGGAATATTGGTTCTGATTTTATTCCTCGGGGCGCCTT		
	TGGAAGGTATACCTGGCTCAAGATATAAAGACGAAGAAAAGAATGGCGTGAAACTG		
	ATCCCAGTAGATCAATTTAAGCCATCTGATGTGGAAATTCAGGCTTGCTTCCGGCAGC		
	AGAACATCGCAGAGCTGTATGGCGCAGTCCTGTGGGTGAACTGTCCATCTCTTTAT		
	GGAAGCAGGCGAGGGAGGGTCTGTTCTGGAGAACTGGAGAGCTGTGGACCAATGAGA		
	GAATTTGAAATTATTTGGGTGACAAAGCATGTTCTCAAGGGACTTGATTTTCTACACT		
	CAAAGAAAGTGATCCATCATGATATTAAACCTAGCAACATTTGTTTTCATGTCCACAAA		
	AGCTGTTTGGTGGATTTTGGCCTAAGTGTCAAATGACCGAAGATGCTATTTTCCCT		
	AAGGACCTCCGAGGAACAGAGATTTACATGAGCCCAGAGGTCATCCTGTGCAGGGGCC		
	ATTCAACCAAAGCAGACATCTACAGCCTGGGGGCCACGCTCATCCACATGACAGCGG		
	CACCCACCCCTGGGTGAAGCGCTACCTCGCTCAGCCTATCCCTCCTACCTGTACATA		
	ATCCACAAGCAAGCACCTCCACTGGAAGACATTGCAGATGACTGCAGTCCAGGGATGA		
	GAGAGCTGATAGAAGCTTCCCTGGAGAGAAACCCCAATCACCGCCCAAGAGCCGCAGA		
	CCTACTAAACATGAGGCCCTGAACCCGCCCAGAGAGGATCAGCCACGCTGTACGAGT		
	CTGGACTCTGCCCTCTTGGAGCGCAAGAGGCTGCTGAGTAGGAAGGAGCTGGAACCTC		
	CTGAGAACATTGGCTGATTCTCTGTGCACAGGAAGCACCGAGGAATCTGAGATGCTCAA		
	GAGGCAACGCTCTCTCTACATCGACCTCGGCGCTCTGGCTGGCTACTTCAATCTTGTT		
	CGGGGACCACCAACGCTTGAATATGGCTGAAGGATGCCATGTTTGCCTCAAATTAAG		
	ACAGCATTGATCTCCTGGAGGCTGGTTCTGCTGCCTCTACACAGGGGCCCGTTACAGT		
	GAATGGTGCCATTTTCAAGGAGCAGTGTGACCTCCTGTGACCCATGAATGTGCCTCC		
	AAGCGGCCCTGTGTGTTTGACATGTGAAGCTATTTGATATGCACCAGGTCTCAAGGTT		
	CTCATTTCTCAGGTGACGTGATTCTAAGGCAGGAATTTGAGAGTTACAGAAGGATCG		
	TGTCTGCTGACTGTTTCATTCACTGTGCACTTTGCTCAAAATTTAAAAATACCAATC		
	ACAAGGATAATAGAGTAGCCTAAAATTACTATTCTTGGTTCTTATTTAAGTATGGAAT		
	ATTCATTTTACTCAGAATAGCCTGTTTGTGTATATTGGTGTATATTATATAACTCTT		
	TGAGCCTTTATTGGTAAATCTGGTATACATTGAATTCATTATAATTTGGGTGACTAG		
	AACAACCTGAAGATTGTAGCAATAAGCTGGACTAGTGTCTCTAAATTTGAGTACATGAT		
	GAATTAGAAGCCATCTGACAGACGGCCACTAGTGACAGTTTCTTTTGTGTTCTATGG		
	AAACATTTTATACTGTACATGCTATGCTGAAGACATTCAAACGTTAGTGTTTTGAATG		
	TGGATAAACTGTGTAAACCACATAATTTGTACATCCAAGGATGAGGTGTGACCTTT		
	AAGAAAAATGAAAACCTTTTGTAATTATTGATGATTTTGTAAATCTTATGACTAAAT		
	TTCTTTTAAGCATTTGTATATTAAAAATAGCATACTGTGTATGTTTATGCAATGCC		
	TTCATGAATCTTTCATACATATATATATTGTAAACATGTAAGTATGTGAGTAGTCTT		
	ATGTAAAGTATGTTTTTACATTATGCAAATAAAACCAATACTTTTGTCCAATGTGGT		
	TGGTCAAATCAACTGAATAAATTCAGTATTTGCCTT		

	ORF Start: ATG at 367		ORF Stop: TGA at 1768
	SEQ ID NO: 2	467 aa	MW at 52896.9kD
NOV1a, CG101683-01 Protein Sequence	MEYMSTGSDNKEEIDLLIKHLNVSDVIDIMENLYASEEPAVYEPSLMTMCQDSNQND RSKSLLLSGQEVPLSSVRYGTVEDLLAFANHISNTAKHFGQRPQESGILLNMVITP QNGRYQIDSDVLLIPWKLTYRNIGSDFIPRGAFGKVYLAQDIKTKKRMACKLI PVDQF KPSDVEIQACFRHENIAELYGAVLWGETVHLFMEAGEGGSVLEKLESCGPMREFEIIW VTKHVLKGLDFLHSHKVIHHDIKPSNIVFMSTKAVLVDFGLSVQMTEDVYFPKDLRG EIIYMSPEVILCRGHSTKADIYSLGATLIHMOTGTPPWVKRYPRSAVPSYLYIIHKQAP PLEDIADDCSPGMRELIEASLERNPNHRPRAADLLKHEALNPPREDQPRCTSLDSALL ERKRLLSRKELELPENIADSSCTGSTBESEMLKRQRSIYIDLALAGYFNLVRGPPTL EYG		
	SEQ ID NO: 3	1425 bp	
NOV1b, 248490507 DNA Sequence	ACCATGGAGTACATGAGCACTGGAAGTGACAATAAAGAAGAGATTGATTTATTAATTA AACATTTAAATGTGTCTGATGTAATAGACATTATGGAAAATCTTTATGCAAGTGAAGA GCCAGCAGTTTATGAACCCAGTCTAATGACCATGTGTCAAGACAGTAATCAAAACGAT GAGCGTTCTAAGTCTCTGCTGCTTAGTGGCCAAGAGGTACCATGGTTGTCTCATCAGTCA GATACGGAAGTGTGGAGGATTGTCTTGTCTTTGCAAACCATATATCAACACTGCAAAA GCATTTTTATGGACAACGACCACAGGAATCTGGAATTTTATTAACATGGTCATCACT CCCCAAAATGGACGTTACCAAATAGATTCCGATGTCTCTCTGATCCCCGGAAGCTGA CTTACAGGAATATTGGTTCTGATTTTATTCCTCGGGGCGCCTTTGGAAAGGTATACTT GGCACAAGATATAAAGACGAAGAAAAGAATGGCGTGTAACTGATCCAGTAGATCAA TTTAAGCCATCTGATGTGGAATCCAGGCTTGCTTCCGGCAGAGAACATCGCAGAGC TGTATGGCGCAGTCTGTGGGGTGAAGTGTCCATCTCTTTATGGAAGCAGGGCGAGGG AGGGTCTGTCTGGAGAACTGGAGAGCTGTGGACCAATGAGAGAATTTGAAATTATT TGGGTGACAAAGCATGTTCTCAAGGGACTTGATTTCTACACTCAAAGAAAGTGATCC ATCATGATATTAAACCTAGCAACATTGTTTTCATGTCCACAAAAGCTGTTTTGGTGGA TTTTGGCCTAAGTGTTCAAATGACCGAAGATGTCTATTTCTTAAGGACCTCCGAGGA ACAGAGATTTACATGAGCCAGAGGTCATCCTGTGCAGGGGCCATTCAACCAAAGCAG ACATCTACAGCCTGGGGGCCAGCTCATCCACATGCAGACGGGCACCCACCTGGGT GAAGCGCTACCTCGCTCAGCCTATCCCTCCTACCTGTACATAATCCACAAGCAAGCA CCTCCACTGGAAGACATTGCAGATGACTGCAGTCCAGGGATGAGAGAGCTGATAGAAG CTTCCCTGGAGAGAAACCCCAATCACCGCCAAGAGCCGAGACCTACTAAACATGA GGCCCTGAACCCGCCAGAGAGGATCAGCCACGCTGTGAGAGTCTGGACTCTGCCCTC TTGGAGCGCAAGAGGCTGCTGAGTAGGAAGGAGCTGGAACCTCTGAGAACATTGCTG ATTCTTCGTGCACAGGAAGCACCAGGAATCTGAGATGCTCAAGAGGCAACGCTCTCT CTACATCGACCTCGGCGCTCTGGCTGGCTACTTCAATCTTGTTCGGGGACCACCAACG CTTGAATATGGCCATCATCACCACCATCACTGA		
	ORF Start: at 1		ORF Stop: TGA at 1423
	SEQ ID NO: 4	474 aa	MW at 53847.9kD
NOV1b, 248490507 Protein Sequence	TMEYMSTGSDNKEEIDLLIKHLNVSDVIDIMENLYASEEPAVYEPSLMTMCQDSNQND ERSKSLLLSGQEVPLSSVRYGTVEDLLAFANHISNTAKHFGQRPQESGILLNMVIT PQNGRYQIDSDVLLIPWKLTYRNIGSDFIPRGAFGKVYLAQDIKTKKRMACKLI PVDQ FKPSDVEIQACFRHENIAELYGAVLWGETVHLFMEAGEGGSVLEKLESCGPMREFEII WVTKHVLKGLDFLHSHKVIHHDIKPSNIVFMSTKAVLVDFGLSVQMTEDVYFPKDLRG TEIYMSPEVILCRGHSTKADIYSLGATLIHMOTGTPPWVKRYPRSAVPSYLYIIHKQA PPLEDIADDCSPGMRELIEASLERNPNHRPRAADLLKHEALNPPREDQPRCQSLDSAL LERKRLLSRKELELPENIADSSCTGSTEESEMLKRQRSIYIDLALAGYFNLVRGPPT LEYGHHHHHH		
	SEQ ID NO: 5	1316 bp	
NOV1c, 253174293 DNA Sequence	ACGGGATCCACCATGGAGTACATGAGCACTGGAAGTGACAATAAAGAAGAGATTGATT TATTAATTAAACATTAAATGTGTCTGATGTAATAGACATTATGGAAAATCTTTATGC AAGTGAAGAGCCAGCAGTTTATGAACCCAGTCTAATGACCATGTGTCAAGACAGTAAT CAAAACGATGAGCGTTCTAAGTCTCTGCTGCTTAGTGGCCAAGAGGTACCATGGTTGT CATCAGTCAGATACGGAAGTGTGAGGATTGCTTGTCTTTGCAAACCATATATCCAA CACTGCAAAGCATTTTTATGGACAACGACCACAGGAATCTGGAATTTTATTAACATG GTCATCACTCCCCAAATGGACGTTACCAAATAGATTCCGATGTTCTCTGATCCCCCT		

	GGAAGCTGACTTACAGGAATATTGGTTCTGATTTTATTCCTCGGGGCGCCTTTGGAAA GGTATACTTGGCACAAGATATAAAGACGAAGAAAAGAATGGCGTGTAACCTGATCCCA GTAGATCAATTTAAGCCATCTGATGTGGAATCCAGGCTTGCTTCCGGCAGGAGAACA TCGCAGAGCTGTATGGCGCAGTCCTGTGGGGTGAACTGTCCATCTCTTTATGGAAGC AGGCGAGGGAGGGTCTGTCTGGAGAACTGGAGAGCTGTGGACCAATGAGAGAATTT GAAATTATTTGGGTGACAAAGCATGTTCTCAAGGGACTTGATTTTCTACACTCAAAGA AAGTGATCCATCATGATATTAACCTAGCAACATTGTTTTCATGTCCACAAAAGCTGT TTTGGTGGATTTTGGCCTAAGTGTTCAAATGACCGAAGATGTCTATTTTCTAAGGAC CTCCGAGGAACAGAGATTTACATGAGCCAGAGGTCATCCTGTGCAGGGGCCATTCAA CCAAAGCAGACATCTACAGCCTGGGGGCCACGCTCATCCACATGCAGACGGGCACCCC ACCCTGGGTGAAGCGCTACCCTCGCTCAGCCTATCCCTCCTACCTGTACATAATCCAC AAGCAAGCACCTCCACTGGAAGACATTGCAGATGACTGCAGTCCAGGGATGAGAGAGC TGATAGAAGCTTCCCTGGAGAGAAACCCCAATCACCGCCCAAGAGCCGAGACCTACT AAAACATGAGGCCCTGAACCCGCCAGAGAGGATCAGCCACCGTGTCTCAGAGCTGGAC TCTGCCCTCTTGGAGCGCAAGAGGCTGCTGAGTAGGAAGGAGCTGGAACCTCTGAGA ACATTGCTCATCATCACCACCATCACTGAGCGGCCGCAAG		
	ORF Start: at 1		ORF Stop: TGA at 1303
	SEQ ID NO: 6	434 aa	MW at 49384.9kD
NOV1c, 253174293 Protein Sequence	TGSTMEYMSTGSDNKEEIDLLIKHLNVSDVIDIMENLYASEEPVAVYEPSLMTMCQDSN QNDERSKSLLSGQEVFWLSSVRYGTVEDLLAFANHISNTAKHFYQRPQESGILLNM VITPQNGRYQIDSDVLLIPWKLTYRNIGSDFIPRGAFGKVYLAQDIKTKRMACKLIP VDQFKPSDVEIQACFRHENIAELYGAVLWGETVHLMFMEAGEGGSVLEKLESCGPMREF EIIWVTKHVLKGLDFLHKKVIHHDIKPSNIVFMSTKAVLVDFGLSVQMTEDVYFPKD LRGTEIYMSPEVILCRGHSTKADIYSLGATLIHMQTGTPPWVKRYPRSAVPSYLIH KQAPPLEDIADDCSPGMRELIEASLERNPNHRPRAADLLKHEALNPREDQPRCQSLD SALLERKRLLSRKELELPENIAHHHHHH		
	SEQ ID NO: 7	1407 bp	
NOV1d, 248490584 DNA Sequence	ACCATGGAGTACATGAGCACTGGAAGTGACAATAAAGAAGAGATTGATTTATTAATTA AACATTTAAATGTGTCTGATGTAATAGACATTATGGAAAATCTTTATGCAAGTGAAGA GCCAGCAGTTTATGAACCCAGTCTAATGACCATGTGTCAAGACAGTAATCAAAACGAT GAGCGTTCTAAGTCTCTGCTGCTTAGTGGCCAAGAGGTACCATGGTTGTCTCATCAGTCA GATACGGAACGTGGAGGATTGCTTGCTTTTGCAAACCATATATCCAACACTGCAAAA GCATTTTATGGACAACGACCACAGGAATCTGGAATTTTATTAACATGGTTCATCACT CCCCAAAATGGACGTTACCAAATAGATTCCGATGTTCTCCTGATCCCTGGAAGCTGA CTTACAGGAATATGGTTCTGATTTTATTCCTCGGGGCGCCTTTGAAAAGGTATACTT GGCACAAGATATAAAGACGAAGAAAAGAATGGCGTGTAACCTGATCCAGTAGATCAA TTTAAGCCATCTGATGTGGAATCCAGGCTTGCTTCCGGCAGCAGAACATCGCAGAGC TGATGGCGCAGTCCGTGGGGTGAACTGTCCATCTCTTTATGGAAGCAGGCGAGGG AGGCTCTGTTCTGGAGAACTGGAGAGCTGTGGACCAATGAGAGAATTGAAATTATT TGGGTGACAAAGCATGTTCTCAAGGGACTTGATTTTCTACACTCAAAGAAAGTGATCC ATCATGATATTAACCTAGCAACATTGTTTTCATGTCCACAAAAGCTGTTTTGGTGGA TTTTGGCCTAAGTGTTCAAATGACCGAAGATGTCTATTTTCTAAGGACCTCCGAGGA ACAGAGATTACATGAGCCAGAGGTCATCCTGTGCAGGGGCCATTCAACCAAAGCAG ACATCTACAGCCTGGGGGCCACGCTCATCCACATGCAGACGGGCACCCACCCCTGGGT GAAGCGCTACCCTCGCTCAGCCTATCCCTCCTACCTGTACATAATCCACAAGCAAGCA CCTCCACTGGAAGACATTGCAGATGACTGCAGTCCAGGGATGAGAGAGCTGATAGAAG CTTCCCTGGAGAGAAACCCCAATCACCGCCCAAGAGCCGAGACCTACTAAAACATGA GGCCCTGAACCCGCCAGAGAGGATCAGCCACGCTGTGAGAGTCTGGACTCTGCCCTC TTGGAGCGCAAGAGGCTGCTGAGTAGGAAGGAGCTGGAACCTCCTGAGAACATTGCTG ATTCTTCGTGCACAGGAAGCACCAGGAATCTGAGATGCTCAAGAGGCAACGCTCTCT CTACATCGACCTCGGCGCTCTGGCTGGCTACTCAATCTTGTTCCGGGACCACCAACG CTTGAATATGGCTGA		
	ORF Start: at 1		ORF Stop: TGA at 1405
	SEQ ID NO: 8	468 aa	MW at 53025.0kD
NOV1d, 248490584	TMEYMSTGSDNKEEIDLLIKHLNVSDVIDIMENLYASEEPVAVYEPSLMTMCQDSNQND ERSKSLLSGQEVFWLSSVRYGTVEDLLAFANHISNTAKHFYQRPQESGILLNMVIT PQNGRYQIDSDVLLIPWKLTYRNIGSDFIPRGAFGKVYLAQDIKTKRMACKLIPVDQ		

Protein Sequence	FKPSDVEIQACFRHENIAELYGAVLWGETVHLFMEAGEGGSVLEKLESCGPMREFEII WVTKHVLKGLDFLHSHKKVIHHDIKPSNIVFMSTKAVLVDFGLSVQMTEDVYFPKDLRG TEIYMSPEVILCRGHSTKADIYSLGATLIHMQTGTPPWVKRYPRSAFPSYLYIIHKQA PPLEDIADDCSPGMRELIEASLERNPNHRPRAADLLKHEALNPPREDQPRCQSLDSAL LERKRLLSRKELELPENIADSSCTGSTEESEMLKRQRSLYIDL GALAGYFNLVRGPPT LEYG		
	SEQ ID NO: 9	1448 bp	
NOVIe, 258054391 DNA Sequence	ACGGGATCCACCATGGGACATCATCACCACCATCACGAGTACATGAGCACTGGAAGTG ACAATAAAGAAGAGATTGATTTATTAATTAAACATTTAAATGTGTCTGATGTAATAGA CATTATGGAAAATCTTTATGCAAGTGAAGAGCCAGCAGTTTATGAACCCAGTCTAATG ACCATGTGTCAAGACAGTAATCAAAACGATGAGCGTTCTAAGTCTCTGCTGCTTAGTG GCCAAGAGGTACCATGGTTGTCTCATCAGTCAGATACGGAACGTGTGGAGGATTGCTTGC TTTGTCAAACCATATATCCAACTGCAAGCATTTTATGGACACGACCACAGGAA TCTGGAATTTATTAACATGGTTCATCACTCCCCAAAATGGACGTTACCAAATAGATT CCGATGTTCTCTGATCCCTGGAAGCTGACTTACAGGAATATTGGTTCTGATTTTAT TCCTCGGGGCGCCTTTTGAAAGGTATACCTTGGCACAGATATAAGACGAAGAAAAGA ATGGCGTGTAACCTGATCCAGTAGATCAATTTAAGCCATCTGATGTGGAATCCAGG CTTGCTTCCGGCACGAGAACATCGCAGAGCTGTATGGCGCAGTCCTGTGGGGTGAAAC TGTCCATCTCTTTATGGAAGCAGGCGAGGGAGGGTCTGTTCTGGAGAACTCGAGAGC TGTGGACCAATGAGAGAATTTGAAATTATTTGGGTGACAAAGCATGTTCTCAAGGGAC TTGATTTTCTACACTCAAAGAAAGTGATCCATCATGATATTAACCTAGCAACATTGT TTTCATGTCCACAAAGCTGTTTTGGTGGATTTTGGCCTAAGTGTTCAAATGACCGAA GATGCTATTTTCTAAGGACCTCCGAGGAACAGAGATTTACATGAGCCCAGAGGTCA TCCTGTGCAGGGGCCATTCAACCAAAGCAGACATCTACAGCCTGGGGGCCACGCTCAT CCACATGCAGACGGGCACCCACCTGGGTGAAGCGCTACCTCGCTCAGCCTATCCC TCCTACCTGTACATAATCCACAAGCAAGCACCTCCACTGGAAGCATTTGCAGATGACT GCAGTCCAGGGATGAGAGAGCTGATAGAAGCTTCCTGGAGAGAAACCCCAATCACCG CCCAAGAGCCGCAGACCTACTAAAACATGAGGCCCTGAACCCGCCAGAGAGGATCAG CCACGCTGTCAAGTCTGGACTCTGCCCTCTTGGAGCGCAAGAGGCTGCTGAGTAGGA AGGAGCTGGAACCTTCCTGAGAACATTGCTGATTCTTCGTGCACAGGAAGCACCGAGGA ATCTGAGATGCTCAAGAGGCAACGCTCTCTTACATCGACCTCGGCGCTCTGGCTGGC TACTTCAATCTTGTTTCGGGGACCACCAACGCTTGAATATGGCTGAGCGGCCGCAAG		
	ORF Start: at 1		ORF Stop: TGA at 1435
	SEQ ID NO: 10	478 aa	MW at 54150.2kD
NOVIe, 258054391 Protein Sequence	TGSTMGHHHHHHEYMSTGSDNKEEIDLLIKHLNVSDVIDIMENLYASEEPAVYEPSLM TMCQDSNQNDERSKSLLSGQEVPLSSVRYGTVEDLLAFANHSNTAKHFGYQRPQE SGILLNMVITPQNGRYQIDSDVLLIPWKLTYRNI GSDFI PRGAFGKVYLAQDIKTKKR MACKLI PVDQFKPSDVEIQACFRHENIAELYGAVLWGETVHLFMEAGEGGSVLEKLES CGPMREFEIIWVTKHVLKGLDFLHSHKKVIHHDIKPSNIVFMSTKAVLVDFGLSVQMT DVYFPKDLRGTEIYMSPEVILCRGHSTKADIYSLGATLIHMQTGTPPWVKRYPRSAFP SYLYIIHKQAPPLEDIADDCSPGMRELIEASLERNPNHRPRAADLLKHEALNPPREDQ PRCQSLDSALLERKRLLSRKELELPENIADSSCTGSTEESEMLKRQRSLYIDL GALAG YFNLVRGPPTLEYG		
	SEQ ID NO: 11	1278 bp	
NOVI f, 248494549 DNA Sequence	ACCATGGAGTACATGAGCACTGGAAGTGACAATAAAGAAGAGATTGATTATTAATTA AACATTTAAATGTGTCTGATGTAATAGACATTATGGAAAATCTTTATGCAAGTGAAGA GCCAGCAGTTTATGAACCCAGTCTAATGACCATGTGTCAAGACAGTAATCAAAACGAT GAGCGTTCTAAGTCTCTGCTGCTTAGTGGCCAAAGAGGTACCATGGTTGTCTCATCAGTCA GATACGGAACGTGGAGGATTGCTTGTCTTTTGCAAACCATATATCCAACTGCAAAA GCATTTTATGGACAACGACCACAGGAATCTGGAATTTTATTAACATGGTTCATCACT CCCCAAAATGGACGTTACCAAATAGATTCCGATGTTCTCTGATCCCTGGAAGCTGA CTTACAGGAATATTGGTTCTGATTTTATTCCTCGGGGCGCCTTTGGAAGGTATACCT GGCACAAGATATAAAGACGAAGAAAAGAAATGGCGTGTAACCTGATCCAGTAGATCAA TTTAAGCCATCTGATGTGGAATCCAGGCTTGCTTCCGGCACGAGAACATCGCAGAGC TGTATGGCGCAGTCTGTGGGGTGAAACTGTCCATCTCTTATGGAAGCAGGCGAGGG AGGGTCTGTTCTGGAGAACTGGAGAGCTGTGGACCAATGAGAGAATTTGAAATTTAT TGGGTGACAAAGCATGTTCTCAAGGGACTTGATTTTCTACACTCAAAGAAAGTGATCC		

	ATCATGATATTAAACCTAGCAACATTGTTTTCATGTCCACAAAAGCTGTTTGGTGGATTTTGGCCTAAGTGTTCAAATGACCGAAGATGTCTATTTTCTAAGGACCTCCGAGGACACAGAGATTTACATGAGCCCAGAGGTCATCCTGTGCAGGGGCCATTCAACCAAAGCAGACATCTACAGCCTGGGGGCCACGCTCATCCACATGCAGACGGGCACCCACCCCTGGGTGAAGCGCTACCCTCGCTCAGCCTATCCCTCCTACCTGTACATAATCCACAAGCAAGCACTCCACTGGAAGACATTGCAGATGACTGCAGTCCAGGGATGAGAGAGCTGATAGAAGCTTCCCTGGAGAGAAACCCCAATCACCGCCCAAGAGCCGACCTACTAAAACATGAGGCCCTGAACCCGCCAGAGAGGATCAGCCACGCTGTGAGAGTCTGGACTCTGCCCTCTGGAGCGCAAGAGGCTGCTGAGTAGGAAGGAGCTGGAACCTCCTGAGAACATTGCTTGA		
	ORF Start: at 1		ORF Stop: TGA at 1276
	SEQ ID NO: 12	425 aa	MW at 48316.8kD
NOV1f, 248494549 Protein Sequence	TMEYMSTGSDNKEEIDLLIKHLNVSDVIDIMENLYASEEPVYEPSLMTMCQDSNQNDERSKSLLLSGQEVFWLSSVRYGTVEDLLAFANHISNTAKHFGYQRPQESGILLNMVITPQNGRYQIDSDVLLIPWKLYRNIQSDFIPRGAFGKVYLAQDIKTKKRMACKLI PVDQFKPSDVEIQACFRHENIAELYGAVLWGETVHLFMEAGEGGSVLEKLESCGPMREFEIIWVTKHVLKGLDFLHKKVHHDIKPSNIVFMSTKAVLVDFGLSVQMTEDVYFPKDLRGTEIYMSPEVILCRGHSTKADIYSLGATLIHMQTGTPPWVKRYPRSAYPSYLYIIHKQAPPLEDIADDCSPGMRELIEASLERNPNHRPRAADLLKHEALNPPREDQPRCQSLDSALLERKRLLSRKELELPENIA		
	SEQ ID NO: 13	1327 bp	
NOV1g, 259741837 DNA Sequence	CCACCATCGGGCGCGGATCCACCATGGGACATCATCACCACCATCAGGAGTACATGAGCACTGGAAGTGACAATAAAGAAGAGATTGATTATTAATTAAACATTTAAATGTGTCTGATGTAATAGACATTATGGAAAATCTTTATGCAAGTGAAGAGCCAGCAGTTTATGAACCCAGTCTAATGACCATGTGTCAAGACAGTAATCAAACGATGAGCGTTCTAAGTCTCTGCTGCTTAGTGGCCAAGAGGTACCATGGTTGTTCATCAGTCAGATACGGAACCTGTGGAGATTGTGCTTGTGTTGCAAACCATATATCCAACACTGCAAAGCATTTTATGGACAACGACCACAGGAATCTGGAATTTTATTAACATGGTCATCACTCCCCAAAATGGACGTTACCAAATAGATTCCGATGTTCTCCTGATCCCTGGAAGCTGACTTACAGGAATATTGGTTCTGATTTTATTCCTCGGGCGCCTTTGGAAAGGTATACTTGGCACAAGATATAAAGACGAAGAAAAGAATGGCGTGTAAACTGATCCCAGTAGATCAATTTAAGCCATCTGATGTGGAAATCCAGGCTTGCTTCCGGCACGAGAACATCGCAGAGCTGTATGGCGCAGTCCCTGTGGGTGAAACTGTCCATCTCTTTATGGAAGCAGGCGAGGGAGGTCTGTTCTGGAGAACTGGAGAGCTGTGGACCAATGAGAGAATTTGAAATTTATTTGGGTGACAAAGCATGTCTCAAGGGACTTGATTTTCTACACTCAAAGAAAGTGATCCATCATGATATTAAACCTAGCAACATTGTTTTTCATGTCCACAAAAGCTGTTTTGGTGGATTTTGGCCTAAGTGTTCAAATGACCGAAGATGTCTATTTTCTAAGGACCTCCGAGGAACAGAGATTTTACATGAGCCCAGAGGTCACTCTGTGCAGGGGCCATTCAACCAAAGCAGACATCTACAGCCTGGGGGCCACGCTCATCCACATGCAGACGGGCACCCACCTGGGTGAAGCGCTACCCCTCGCTCAGCCTATCCCTCCTACCTGTACATAATCCACAAGCAAGCACCTCCACTGGAAGACATGTCAGATGACTGCAGTCCAGGGATGAGAGAGCTGATAGAAGCTTCCCTGGAGAGAAACCCAATCACCGCCCAAGAGCCGACGCTACTAAAACATGAGGCCCTGAACCCGCCAAGAGAGGATCAGCCACGCTGTGAGAGTCTGGACTCTGCCCTCTTGAGAGCGCAAGAGGCTGCTGAGTAGGAAGGAGCTGGAACCTCCTGAGAACATTGCTTGAGGCGGCCG		
	ORF Start: at 3		ORF Stop: TGA at 1317
	SEQ ID NO: 14	438 aa	MW at 49768.4kD
NOV1g, 259741837 Protein Sequence	TIGRGSTMGHHHHHHEYMSTGSDNKEEIDLLIKHLNVSDVIDIMENLYASEEPVYEPSLMTMCQDSNQNDERSKSLLLSGQEVFWLSSVRYGTVEDLLAFANHISNTAKHFGYQRPQESGILLNMVITPQNGRYQIDSDVLLIPWKLYRNIQSDFIPRGAFGKVYLAQDIKTKKRMACKLI PVDQFKPSDVEIQACFRHENIAELYGAVLWGETVHLFMEAGEGGSVLEKLESCGPMREFEIIWVTKHVLKGLDFLHKKVHHDIKPSNIVFMSTKAVLVDFGLSVQMTEDVYFPKDLRGTEIYMSPEVILCRGHSTKADIYSLGATLIHMQTGTPPWVKRYPRSAYPSYLYIIHKQAPPLEDIADDCSPGMRELIEASLERNPNHRPRAADLLKHEALNPPREDQPRCQSLDSALLERKRLLSRKELELPENIA		
	SEQ ID NO: 15	1428 bp	
NOV1h,	ACCATGGGACATCATCACCACCATCAGGAGTACATGAGCACTGGAAGTGACAATAAAG		

260480803 DNA Sequence	AAGAGATTGATTTATTAATTAACATTTAAATGTGTCTGATGTAATAGACATTATGGA AAATCTTTATGCAAGTGAAGAGCCAGCAGTTTATGAACCCAGTCTAATGACCATGTGT CAAGACAGTAATCAAAACGATGAGCGTTCTAAGTCTCTGCTTAGTGGCCAAGAGG TACCATGGTTGTCTATCAGTCAGATACGGAACGTGGAGGATTGCTTGCTTTTGCAAA CCATATATCCAACACTGCAAAGCATTTTATGGACAACGACCACAGGAATCTGGAATT TTATTAACATGGTCATCACTCCCAAAATGGACGTTACCAAATAGATTCCGATGTTT TCCTGATCCCTGGAAGCTGACTTACAGGAATATTGGTTCTGATTTTATTCCTCGGGG CGCCTTTGGAAAGGTATACTTGGCACAAGATATAAAGACGAAGAAAAGAATGGCGTGT AAACTGATCCAGTAGATCAATTTAAGCCATCTGATGTGGAAATCCAGGCTTGCTTCC GGCAGGAGAACATCGCAGAGCTGTATGGCGCAGTCTGTGGGGTGAAACTGTCCATCT CTTTATGGAAGCAGGCGAGGGAGGGTCTGTTCTGGAGAACTGGAGAGCTGTGGACCA ATGAGAGAATTTGAAATTATTTGGGTGACAAAGCATGTTCTCAAGGGACTTGATTTTC TACACTCAAAGAAAGTGATCCATCATGATATTAAACCTAGCAACATTGTTTTCATGTC CACAAAGCTGTTTTGGTGGATTTTGGCCTAAGTGTTCAAATGACCGAAGATGCTCTAT TTTCTTAAGGACCTCCGAGGAACAGAGATTTACATGAGCCCAGAGGTATCCTGTGCA GGGGCCATTCAACCAAAGCAGACATCTACAGCCTGGGGGCCACGCTCATCCACATGCA GACGGGCACCCACCTGGGTGAAGCGCTACCTCGCTCAGCCTATCCCTCCTACCTG TACATAATCCACAAGCAAGCACCTCCACTGGAAGACATTGCAGATGACTGCAGTCCAG GGATGAGAGAGCTGATAGAAGCTTCCCTGGAGAGAAACCCAATCACCGCCCAAGAGC CGCAGACCTACTAAACATGAGGCCCTGAACCCGCCAGAGAGGATCAGCCACGCTGT CAGAGTCTGGAATCTGCCCTCTTGGAGCGCAAGAGGCTGCTGAGTAGGAAGGAGCTGG AACTTCTGAGAACATTGCTGATTCTTCGTGCACAGGAAGCACCAGGAATCTGAGAT GCTCAAGAGGCAACGCTCTCTACATCGACCTCGGCGCTCTGGCTGGCTACTTCAAT CTTGTTCGGGGACCACCAACGCTTGAATATGGCTGA		
	ORF Start: at 1		ORF Stop: TGA at 1426
	SEQ ID NO: 16	475 aa	MW at 53904.9kD
NOV1h, 260480803 Protein Sequence	TMGHHHHHHEYMSTGSDNKEEIDLLIKHLNVSDVIDIMENLYASEEPAVYEPSLMTMC QDSNQNDERSKSLLLSGQEVFWLSSVRYGTVEDLLAFANHISNTAKHFGQRPQESGI LLNMVITPQNGRYQIDSDVLLIPWKLYRNIGSDFIPRGAFGKVYLAQDIKTKRMAC KLIPVDQFKPSDVEIQACFRHENIAELYGAVLWGETVHLFMEAGEGSSVLEKLESCGP MREFEIIWVTKHVLKGLDFLHSSKKVIHHDIKPSNIVFMSTKAVLVDFGLSVQMTEDVY FPKDLRGTEIYMSPEVILCRGHSTKADIYSLGATLIHMOTGTPPWVKRYPRSAYSYL YIIHKQAPPLEDIADDCSPGMRELI EASLERNPNHRPRAADLLKHEALNPREDQPRC QSLDSALLERKRLLSRKELELPENIADSSCTGSTEESEMLKRQRSLYIDL GALAGYFN LVRGPPTLEYG		
	SEQ ID NO: 17	1434 bp	
NOV1i, 209983329 DNA Sequence	CGCGGATCCACCATGGAGTACATGAGCACTGGAAGTGACAATAAAGAAGAGATTGATT TATTAATTAACATTTAAATGTGTCTGATGTAATAGACATTATGGAAAATCTTTATGC AAGTGAAGAGCCAGCAGTTTATGAACCCAGTCTAATGACCATGTGTCAAGACAGTAAT CAAAACGATGAGCGTTCTAAGTCTCTGCTGCTTAGTGGCCAAGAGGTACCATGTTGT CATCAGTCAGATACGGAACGTGGAGGATTGCTTGCTTTTGCAAACCATATATCCAA CACTGCAAAGCATTTTTATGGACAACGACCACAGGAATCTGGAATTTTATTAACATG GTCATCACTCCCAAAATGGACGTTACCAAATAGATTCCGATGTTCTCCTGATCCCT GGAAGCTGACTTACAGGAATATTGGTTCTGATTTTATTCCTCGGGGCGCCTTTGAAA GGTATACTTGGCACAAGATATAAAGACGAAGAAAAGAATGGCGTGTAACATGATCCCA GTAGATCAATTTAAGCCATCTGATGTGGAATCCAGGCTTGCTTCCGGCAGCAGAAACA TCGCAGAGCTGTATGGCGCAGTCTGTGGGGTGAAACTGTCCATCTCTTTATGGAAGC AGGCGAGGGAGGGTCTGTTCTGGAGAACTGGAGAGCTGTGGACCAATGAGAGAATTT GAAATTATTTGGGTGACAAAGCATGTTCTCAAGGGACTTGATTTTCTACACTCAAAGA AAGTGATCCATCATGATATTAAACCTAGCAACATTGTTTTCATGTCCACAAAAGCTGT TTTGGTGGATTTTGGCCTAAGTGTTCAAATGACCGAAGATGTCTATTTTCTAAGGAC CTCCGAGGAACAGAGATTTACATGAGCCCAGAGGTATCCTGTGCAGGGGCCATTCAA CCAAAGCAGACATCTACAGCCTGGGGGCCACGCTCATCCACATGCAGACGGGCACCCC ACCTGGGTGAAGCGTACCTCGCTCAGCCTATCCCTCCTACCTGTACATAATCCAC AAGCAAGCACCTCCACTGGAAGACATTGCAGATGACTGCAGTCCAGGGATGAGAGAGC TGATAGAAGCTTCCCTGGAGAGAAACCCCAATCACCGCCCAAGAGCGCAGACCTACT AAAACATGAGGCCCTGAACCCGCCAGAGAGGATCAGCCACGCTGTGAGACTCTGGAC TCTGCCCTCTTGGAGCGCAAGAGGCTGCTGAGTAGGAAGGAGCTGGAATCTCTGAGA		

	ACATTGCTGATTCTTCGTGCACAGGAAGCACCGAGGAATCTGAGATGCTCAAGAGGCA ACGCTCTCTCTACATCGACCTCGGCGCTCTGGCTGGCTACTTCAATCTTGTTCGGGGA CCACCAACGCTTGAATATGGCTGAGCGGCCGCTTTTTCCTT		
	ORF Start: at 1		ORF Stop: TGA at 1414
	SEQ ID NO: 18	471 aa	MW at 53325.3kD
NOV1i, 209983329 Protein Sequence	RGSTMEYMSTGSDNKEEIDLLIKHLNVSDVIDIMENLYASEEPAVYEPSLMTMCQDSN QNDERSKSLLLSGQEVFWLSSVRYGTVEDLLAFANHISNTAKHFGQRPQESGILLNM VITPQNGRYQIDSDVLLIPWKLTYRNIGSDFIPRGAFGKVYLAQDIKTKRMACKLIP VDQFKPSDVEIQACFRHENIAELYGAVLWGETVHLFMEAGEGGSVLEKLESCGPMREF EIIWVTKHVLKGLDFLHSHKVIHHDIKPSNIVFMSTKAVLVDFGLSVQMTEDVYFPKD LRGTEIYMSPEVILCRGHSTKADIYSLGATLIHMQTGTPPWVKRYPRSAVPSYLYIIH KQAPPLIEDIADDCSPGMRELIEASLERNPNHRPRAADLLKHEALNPPREDQPRCQSLD SALLERKRLSRKELELPENIADSSCTGSTEESEMLKRQRSYIDL GALAGYFNLVRG PPTLEYG		
	SEQ ID NO: 19	1772 bp	
NOV1j, 212779055 DNA Sequence	TGTCGTAACAACTCCGCCCCATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGG TCTATATAAGCAGAGCTCTCTGGCTAACTAGAGAACCCACTGCTTACTGGCTTATCGA AATTAATACGACTCACTATAGGGAGACCCAAGCTGGCTAGCGTTTAACTTAAGCTTG GTACCGAGCTCGGATCCACCATGGAGTACATGAGCACTGGAAGTGACAATAAAGAAGA GATTGATTTATTAATTAACATTTAAATGTGTCTGATGTAATAGACATTATGGAAAT CTTTATGCAAGTGAAGAGCCAGCAGTTTATGAACCCAGTCTAATGACCATGTGTCAAG ACAGTAATCAAAACGATGAGCGTCTTAAGTCTCTGCTGCTTAGTGGCCAAGAGGTACC ATGGTTGTGTCATCAGTCAGATACGGAACCTGTGGAGGATTGCTTGCTTTTGCAAAACCAT ATATCCAACACTGCAAAGCATTTTTTATGGACAACGACCACAGGAATCTGGAATTTTAT TAAACATGGTCATCACTCCCCAAATGGACGTTACCAAATAGATTCCGATGTTCTCCT GATCCCCCTGGAAGCTGACTTACAGGAATATTGGTTCTGATTTTATCCTCGGGGCGCC TTTGGAAGGTATACCTGGCACAAGATATAAAGACCAAGAAAAGAAATGGCGTGTAAC TGATCCCACTAGATCAATTTAAGCCATCTGATGTGGAATCCAGGCTTGCTTCCGGCA CGAGAACATCGCAGAGCTGTATGGCGCAGTCCTGTGGGTGAAACTGTCCATCTCTTT ATGGAAGCAGGCGAGGAGGGTCTGTTCTGGAGAACTGGAGAGCTGTGGACCAATGA GAGAATTTGAAATTATTTGGGTGACAAAGCATGTTCTCAAGGACTTGATTTTCTACA CTCAAAGAAAGTGATCCATCATGATTAACCTAGCAACATGTGTTTTCATGTCCACA AAAGCTGTTTTGGTGGATTTTGGCCTAAGTGTTCAAATGACCGAAGATGTCTATTTTC CTAAGGACCTCCGAGGAACAGAGATTTACATGAGCCAGAGGTCATCTGTCAGGGG CCATTCAACCAAAGCAGACATCTACAGCCTGGGGGCCACGCTCATCCACATGCAGACG GGCACCACCCCTGGGTGAAGCGCTACCTCGCTCAGCCTATCCCTCCTACCTGTACA TAATCCACAAGCAAGCACCTCCACTGGAAGACATTGCAGATGACTGCAGTCCAGGGAT GAGAGAGCTGATAGAAGCTTCCCTGGAGAGAAACCCCAATCACCGCCCAAGAGCCGCA GACCTACTAAACATGAGGCCCTGAACCCGCCAGAGAGGATCAGCCACGCTGTGAGA GTCTGGACTCTGCCCTCTTGGAGCGCAAGAGGCTGCTGAGTAGGAAGGAGCTGGAAC TCCTGAGAACATTGCTGATTCTTCGTGCACAGGAAGCACCAGGAATCTGAGATGCTC AAGAGGCAACGCTCTCTACATCGACCTCGGCGCTCTGGCTGGCTACTTCAATCTTG TTCGGGGACCACCAACGCTTGAATATGGCTGAGCGGCCGCTCGAGTCTAGAGGGCCCC TTTAAACCCGCTGATCAGCCTCGACTGTGCCCTTCTAGTTGCCAGCCATCTGTTGTTTG CCCCCTCCCCGTCCTTCTTGACCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAA TAAATGAGGAAATTGCATCGCATTGTCTGAG		
	ORF Start: at 138		ORF Stop: TGA at 1596
	SEQ ID NO: 20	486 aa	MW at 54926.2kD
NOV1j, 212779055 Protein Sequence	GDPSWLAFKLLKLTGELGSTMEYMSTGSDNKEEIDLLIKHLNVSDVIDIMENLYASEE AVYEPSLMTMCQDSNQNDERSKSLLLSGQEVFWLSSVRYGTVEDLLAFANHISNTAKH FYGQRPQESGILLNMVITPQNGRYQIDSDVLLIPWKLTYRNIGSDFIPRGAFGKVYLA QDIKTKRMACKLIPVDQFKPSDVEIQACFRHENIAELYGAVLWGETVHLFMEAGEGG SVLEKLESCGPMREFEIIWVTKHVLKGLDFLHSHKVIHHDIKPSNIVFMSTKAVLVDF GLSVQMTEDVYFPKDLRGTEIYMSPEVILCRGHSTKADIYSLGATLIHMQTGTPPWVK RYPRSAVPSYLYIIHKQAPPLIEDIADDCSPGMRELIEASLERNPNHRPRAADLLKHEA LNPPREDQPRCQSLD SALLERKRLSRKELELPENIADSSCTGSTEESEMLKRQRSY IDL GALAGYFNLVRGPPTLEYG		

	SEQ ID NO: 21	1770 bp	
NOV1k, 212779063 DNA Sequence	<p>TTCGTAACAACCTCCGCCCCATTGACGCAAAATGGGCGGTAGGCGTGTACGGTGGGAGGT  CTATATAAGCAGAGCTCTCTGGCTAACTAGAGAACCCACTGCTTACTGGCTTATCGAA  ATTAATACGACTCACTATAGGGAGACCCAAGCTGGCTAGCGTTTAAACTTAAGCTTGG  TACCGAGCTCGGATCCACCATGGAGTACATGAGCACTGGAAGTGACAATAAAGAAGAG  ATTGATTTATTAATTAACATTTAAATGTGTCTGATGTAATAGACATTATGAAAAATC  TTTATGCAAGTGAAGAGCCAGCAGTTTATGAACCCAGTCTAATGACCATGTGTCAAGA  CAGTAATCAAAACGATGAGCGTTCTAAGTCTCTGCTGCTTAGTGGCCAAGAGGTACCA  TGGTTGTCATCAGTCAGATATGGAAGTGTGGAGGATTTGCTTGCTTTTGCAAAACATA  TATCCAACACTGCAAAGCATTTTTATGGACAACGACCACAGGAATCTGGAATTTTATT  AAACATGGTCATCACTCCCCAAAATGGACGTTACCAAATAGATTCCGATGTTCTCCTG  ATCCCCTGGAAGCTGACTTACAGGAATATTGGTTCTGATTTTATTCTCGGGGCGCCT  TTGGAAGGTATACCTGGGCACAAGATATAAAGACGAAGAAAAGAATGGCGTGTAAACT  GATCCCAGTAGATCAATTTAAGCCATCTGATGTGGAATCCAGGCTTGCTTCCGGCAC  GAGAACATCGCAGAGCTGTATGGCGCAGTCCGTGTTGGGTGAACTGTCCATCTCTTTA  TGGAAGCAGGCGAGGGAGGGTCTGTTCTGGAGAACTGGAGAGCTGTGGACCAATGAG  AGAATTTGAAATTATTTGGGTGACAAAGCATGTTCTCAAGGGACTTGATTTTCTACAC  TCAAAGAAAGTGATCCATCATGATATTAACCTAGCAACATTGTTTTCTATGTCACAA  AAGCTGTTTTGGTGGATTTTGGCCTAAGTGTTCAAATGACCGAAGATGTCTATTTTCC  TAAGGACCTCCGAGGAACAGAGATTACATGAGCCCAGAGGTATCTGTGCAGGGGC  CATTCAACCAAAGCAGACATCTACAGCCTGGGGGCCACGCTCATCCACATGCAGACGG  GCACCCACCCCTGGGTGAAGCGCTACCCTCGCTCAGCCTATCCCTCCTACCTGTACAT  AATCCACAAGCAAGCACCTCCACTGGAAGACATGTCAGATGACTGCAGTCCAGGGATG  AGAGAGCTGATAGAAGCTTCCCTGGAGAGAAACCCCAATCACCGCCCAAGAGCCGCAG  ACCTACTAAAACATGAGGCCCTGAACCCGCCCAGAGAGGATCAGCCACGCTGTACAG  TCTGGACTCTGCCCTCTTGGAGCGCAAGAGGCTGCTGAGTAGGAAGGAGCTGGAACCT  CCTGAGAACATTGCTGATTCTTCTGTCACAGGAAGCACCGAGGAATCTGAGATGCTCA  AGAGGCAACGCTCTCTCTACATCGACCTCGGCGCTCTGGCTGGCTACTTCAATCTTGT  TCGGGGACCACCAACGCTTGAATATGGCTGAGCGGCCGCTCGAGTCTAGAGGGCCCCGT  TTAAACCCGCTGATCAGCCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGC  CCCTCCCCCGTGCTTCTTACCCTGGAAGGTGCCACTCCCAGTGTCTTCTCTAAT  AAAATGAGGAAATTGCATCGCATTGTCTGA</p>		
	ORF Start: at 137		ORF Stop: TGA at 1595
	SEQ ID NO: 22	486 aa	MW at 54926.2kD
NOV1k, 212779063 Protein Sequence	<p>GDPSWLAFKLLKLTGELGSTMEYMTGSDNKEEIDLLIKHLNVSDVIDIMENLYASEEP  AVYEPSLMTMCQDSNQNDERSKSLLLSGQEVFWLSSVRYGTVEDLLAFANHISNTAKH  FYGQRPQESGILNMVITPQNGRYQIDSDVLLIPWKLTYRNIIGSDFI PRGAFGKVYLA  QDIKTKKRMACKLIPVDQFKPSDVEIQACFRHENIAELYGAVLWGETVHLFMEAGEGG  SVLEKLESCGPMREFEIIWVTKHVLKGLDFLHKKVIHHDIKPSNIVFMSTKAVLVDF  GLSVQMTEDVYFPKDLRGTEI VMSPEVILCRGHSTKADIYSLGATLIHMQTGTPPWVK  RYPRSAYPYLYI IHKQAPPLIEDIADDCSPGMRELIEASLERNPNHRPRAADLLKHEA  LNPPREDQPRCQLSDSALLERKRLLSRKELELPENIADSSCTGSTEESEMLKRQSLY  IDLGALAGYFNLVRGPPTLEYG</p>		
	SEQ ID NO: 23	1772 bp	
NOV11, CG101683-02 DNA Sequence	<p>TGTCGTAACAACCTCCGCCCCATTGACGCAAAATGGGCGGTAGGCGTGTACGGTGGGAGG  TCTATATAAGCAGAGCTCTCTGGCTAACTAGAGAACCCACTGCTTACTGGCTTATCGA  AATTAATACGACTCACTATAGGGAGACCCAAGCTGGCTAGCGTTTAAACTTAAGCTTG  GTACCGAGCTCGGATCCACCATGGAGTACATGAGCACTGGAAGTGACAATAAAGAAGA  GATTGATTTATTAATTAACATTTAAATGTGTCTGATGTAATAGACATTATGAAAAAT  CTTTATGCAAGTGAAGAGCCAGCAGTTTATGAACCCAGTCTAATGACCATGTGTCAAG  ACAGTAATCAAAACGATGAGCGTTCTAAGTCTCTGCTGCTTAGTGGCCAAGAGGTACC  ATGTTTGTCACTAGTCAGATACGGAAGTGTGGAGGATTTGCTTGCTTTTGCAAACCAT  ATATCCAACACTGCAAAGCATTTTTATGGACAACGACCACAGGAATCTGGAATTTTAT  TAAACATGGTCATCACTCCCCAAAATGGACGTTACCAAATAGATTCCGATGTTCTCCT  GATCCCCTGGAAGCTGACTTACAGGAATATTGGTTCTGATTTTATTCTCGGGGCGCC  TTTGGAAGGTATACCTGGGCACAAGATATAAAGACGAAGAAAAGAATGGCGTGTAAAC  TGATCCCAGTAGATCAATTTAAGCCATCTGATGTGGAATCCAGGCTTGCTTCCGGCA  CGAGAACATCGCAGAGCTGTATGGCGCAGTCCGTGTTGGGTGAACTGTCCATCTCTTT</p>		

	ATGGAAGCAGGCGAGGGAGGGTCTGTTCTGGAGAACTGGAGAGCTGTGGACCAATGA GAGAATTTGAAATTATTTGGGTGACAAAGCATGTTCTCAAGGGACTTGATTTTCTACA CTCAAAGAAAGTGATCCATCATGATATTAACCTAGCAACATTGTTTTCATGTCCACA AAAGCTGTTTGGTGGATTGTCCTAAGTGTTCAAATGACCGAAGATGTCTATTTTC CTAAGGACCTCCGAGGAACAGAGATTACATGAGCCAGAGGTCATCTGTGCAGGGG CCATTCAACCAAAGCAGACATCTACAGCCTGGGGGCCACGCTCATCCACATGCAGACG GGCACCCACCCCTGGGTGAAGCGCTACCCTCGCTCAGCCTATCCCTCCTACCTGTACA TAATCCACAAGCAAGCACCTCCACTGGAAGACATTGCAGATGACTGCAGTCCAGGGAT GAGAGAGCTGATAGAAGCTTCCCTGGAGAGAAACCCAATCACCGCCCAAGAGCCGCA GACCTACTAAAACATGAGGCCCTGAACCCGCCAGAGAGGATCAGCCACGCTGTCAGA GTCTGGACTCTGCCCTCTTGGAGCGCAAGAGGCTGCTGAGTAGGAAGGAGCTGGAAC TCCTGAGAACATTGCTGATTCTTCGTGCACAGGAAGCACCAGGAATCTGAGATGCTC AAGAGGCAACGCTCTCTACATCGACCTCGGCGCTCTGGCTGGCTACTTCAATCTTG TTCGGGGACCAACCGCTTGAATATGGCTGAGCGGCCGCTCGAGTCTGAGAGGGCCCG TTAAACCCGCTGATCAGCCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTG CCCCTCCCGTGCCTTCTTGACCTGGAAGGTGCCACTCCACTGTCTTTCTTAA TAAATGAGGAAATTGCATCGCATTTGCTGAG		
	ORF Start: ATG at 195		ORF Stop: TGA at 1596
	SEQ ID NO: 24	467 aa	MW at 52923.9kD
NOV11, CG101683-02 Protein Sequence	MEYMSGSDNKEEIDLLIKHLNVSDVIDIMENLYASEEPVYEPSLMTMCQDSNQND RSKSLLLSGQEVPLSSVRYGTVEDLLAFANHISNTAKHFGYQRPQESGILLNMVITP QNGRYQIDSDVLLIPWKLTYRNIGSDFIPRGAFGKVYLAQDIKTKRMACKLIPVDQF KPSDVEIQACFRHENIAELYGAVLWGETVHLFMEAGEGGSVLEKLESCGPMREFEIIW VTKHVLKGLDFLHKKVVIHHDIKPSNIVFMSTKAVLVDFGLSVQMTEDVYFPKDLRGT EIYMSPEVILCRGHSTKADIYSLGATLIHMOTGTPPWVKRYPRSAYPSYLYIIHKQAP PLEDIADDCSPGMRELIEASLERNPNHRPRAADLLKHEALNPPREDQPRCQSLDSALL ERKRLLSRKELELPENIADSSCTGSTEESEMLKRQSLYIDLALAGYFNLVRGPPTL EYG		
	SEQ ID NO: 25	1425 bp	
NOV1m, CG101683-03 DNA Sequence	ACCATGGAGTACATGAGCACTGGAAGTGACAATAAAGAAGAGATTGATTTTATTAATTA AACATTTAAATGTGTCTGATGTAATAGACATTATGGAATCTTTATGCAAGTGAAGA GCCAGCAGTTTATGAACCCAGTCTAATGACCATGTGTCAAGCAGTAATCAAAACGAT GAGCGTTCTAAGTCTCTGCTGCTTAGTGGCCAGAGGTACCATGGTTGTTCATCAGTCA GATACGGAACGTGTGGAGGATTGCTTTGCTTTTGCAAACCATATATCAACACTGCAAA GCATTTTATGGACAACGACCACAGGAATCTGGAATTTTATTAACATGGTCATCACT CCCCATAATGGACGTTACCAATAGATTCCGATGTTCTCCTGATCCCCTGGAAGCTGA CTTACAGGAATATTGGTTCTGATTTTATTCTCGGGGCGCCTTTGGAAGGTATACTT GGCACAAGATATAAGACGAAGAAAAGATGGCGTGTAACCTGATCCAGTAGATCAA TTTAAGCCATCTGATGTGGAATCCAGGCTTGCTTCGGGCAGGAGACATCGCAGAGC TGTATGGCGCAGTCTGTGGGGTGAAACTGTCCATCTCTTTATGGAAGCAGGCGAGGG AGGGTCTGTTCTGGAGAACTGGAGAGCTGTGGACCAATGAGAGAATTTGAAATTATT TGGGTGACAAAGCATGTTCTCAAGGGACTTGATTTTCTACACTCAAAGAAAGTGATCC ATCATGATATTAACCTAGCAACATTGTTTTTCATGTCCACAAAAGCTGTTTTGGTGGGA TTTTGGCCTAAGTGTCAAATGACCGAAGATGTCTATTTTCTAAGGACCTCCGAGGA ACAGAGATTTACATGAGCCAGAGGTCTCCTGTGCAGGGGCCATTCAACCAAAGCAG ACATCTACAGCCTGGGGGCCACGCTCATCCACATGCAGACGGGCACCCACCCCTGGGT GAAGCGCTACCTCGCTCAGCCTATCCCTCTACCTGTACATAATCCACAAGCAAGCA CCTCCACTGGAAGACATTGCAGATGACTGCAGTCCAGGGATGAGAGAGCTGATAGAAG CTTCCCTGGAGAGAAACCCAATCACCGCCCAAGAGCCGACCTACTAAAACATGA GGCCCTGAACCCGCCAGAGAGGATCAGCCACGCTGTGCAGAGCTGGAAGCTGCGCTC TTGGAGCGCAAGAGGCTGCTGAGTAGGAAGGAGCTGGAAGCTTCTGAGAACATTGCTG ATTCTTCGTGCACAGGAAGCACCAGGAATCTGAGATGCTCAAGAGGCAACGCTCTCT CTACATCGACCTCGGCGCTCTGGCTGGCTACTTCAATCTGTTTGGGGACCAACG CTTGAATATGGCCATCATCACCACCATCACTGA		
	ORF Start: at 1		ORF Stop: TGA at 1423
	SEQ ID NO: 26	474 aa	MW at 53847.9kD
NOV1m,	TMEYMSGSDNKEEIDLLIKHLNVSDVIDIMENLYASEEPVYEPSLMTMCQDSNQND		

CG101683-03 Protein Sequence	ERSKSLLLSGQEV PWLSSVRYGTVEDLLAFANHISNTAKHFYQORPQESGILLNMVIT PQNGRYQIDSDVLLIPWKLTYRNIGSDFIPRGAFGKVYLAQDIKTKRMACKLI PVDQ FKPSDVEIQACFRHENIAELYGAVLWGETVHFLFMEAGEGGSVLEKLESCGPMREFEII WVTKHVLKGLDFLHKKVIHHDIKPSNIVFMSTKAVLVDFGLSVQMTEDVYFPKDLRG TEIYMSPEVILCRGHSTKADIYSLGATLIHMQTGTPPWVKRYPRSA YPSYLYIIHKQA PPLEDIADDCSPGMRELIEASLERNPNHRPRAADLLKHEALNPREDQPRCQSLDSAL LERKRLLSRKELELPENIADSSCTGSTEESEMLKRQRSLYIDL GALAGYFNLVRGPPT LEYGHHHHHH		
	SEQ ID NO: 27	1344 bp	
NOV1n, CG101683-04 DNA Sequence	ACCATGGAGTACATGAGCACTGGAAGTGACAATAAAGAAGAGATTGATTTATTAATTA AACATTTAAATGTGTCTGATGTAATAGACATTATGAAAAATCTTTATGCAAGTGAAGA GCCAGCAGTTTATGAACCCAGTCTAATGACCATGTGTCAAGACAGTAATCAAAACGAT GAGCGTTTCTAAGTCTCTGCTGCTTAGTGGCCAAGAGGTACCATGGTTGTCATCAGTCA GATACGGAAGTGTGGAGGATTTGCTTGTCTTTTGCAAACCATATATCCAACACTGCAAA GCATTTTTATGGACAACGACCACAGGAATCTGGAATTTTATTAACATGGTCATCACT CCCCAAAATGGACGTTACCAAATAGATTCCGATGTTCTCCTGATCCCCCTGGAAGCTGA CTTACAGGAATATTGGTTCTGATTTTATTCCTCGGGGCGCCTTTGGAAAGGTATACCTT GGCACAAGATATAAAGACGAAGAAAAGAATGGCGTGTAACATGATCCAGTCAACAA TTTAAGCCATCTGATGTGGAATCCAGGCTTGCTTCCGGCAGCAGAAACATCGCAGAGC TGTATGGCGCAGTCTGTGGGGTGAACTGTCCATCTCTTTATGGAAGCAGGCGAGGG AGGGTCTGTCTGGAGAACTGGAGAGCTGTGGACCAATGAGAGAATTTGAAATTATT TGGGTGACAAAGCATGTTCTCAAGGGACTTGATTTTCTACACTCAAAGAAAGTGATCC ATCATGATATTAAACCTAGCAACATGTTTTTCATGTCCACAAAGCTGTTTTGGTGGA TTTTGGCCTAAGTGTTCAAATGACCGAAGATGTCTATTTTCCCTAAGGACCTCCGAGGA ACAGAGATTTACATGAGCCAGAGGTCATCCTGTGCAGGGGCCATTCAACCAAAGCAG ACATCTACAGCCTGGGGGCCACGCTCATCCACATGCAGACGGGCACCCACCTGGGT GAAGCGCTACCCTCGCTCAGCCTATCCCTCCTACCTGTACATAATCCACAAGCAAGCA CCTCCACTGGAAGACATTGCAGATGACTGCAGTCCAGGGATGAGAGAGCTGATAGAAG CTTCCCTGGAGAGAAACCCCAATCACCGCCCAAGAGCCGACAGCTACTAAACATGA GGCCCTGAACCCGCCAGAGAGGATCAGCCACGCTGTGAGTCTGGACTCTGCCCTC TTGGAGCGCAAGAGGCTGCTGAGTAGGAAGGAGCTGGAACCTCTGAGAACATTGCTC ATCATCACCACCATCACTGAGCGGCCGCTTCGATCTAGAGTGCAGTCTCGAGCATG CGGTACCAGC		
	ORF Start: at 1		ORF Stop: TGA at 1294
	SEQ ID NO: 28	431 aa	MW at 49139.7kD
NOV1n, CG101683-04 Protein Sequence	TMEYMSTGSDNKEEIDLLIKHLNVSDVIDIMENLYASEEPAVYEPSLMTMCQDSNQND ERSKSLLLSGQEV PWLSSVRYGTVEDLLAFANHISNTAKHFYQORPQESGILLNMVIT PQNGRYQIDSDVLLIPWKLTYRNIGSDFIPRGAFGKVYLAQDIKTKRMACKLI PVDQ FKPSDVEIQACFRHENIAELYGAVLWGETVHFLFMEAGEGGSVLEKLESCGPMREFEII WVTKHVLKGLDFLHKKVIHHDIKPSNIVFMSTKAVLVDFGLSVQMTEDVYFPKDLRG TEIYMSPEVILCRGHSTKADIYSLGATLIHMQTGTPPWVKRYPRSA YPSYLYIIHKQA PPLEDIADDCSPGMRELIEASLERNPNHRPRAADLLKHEALNPREDQPRCQSLDSAL LERKRLLSRKELELPENIAHHHHHH		
	SEQ ID NO: 29	1327 bp	
NOV1o, CG101683-05 DNA Sequence	CCACCATCGGGCGCGGATCCACCATGGGACATCATCACCACCATCAGAGTACATGAG CACTGGAAGTGACAATAAAGAAGAGATTGATTATTAATTAACATTTAAATGTGTCT GATGTAATAGACATTATGAAAAATCTTTATGCAAGTGAAGAGCCAGCAGTTTATGAAC CCAGTCTAATGACCATGTGTCAAGACAGTAATCAAAACGATGAGCGTTCTAAGTCTCT GCTGCTTAGTGGCCAAGAGGTACCATGGTTGTCATCAGTCAGATACGGAAGTGTGGAG GATTTGCTTGCTTTTGCAAACCATATATCCAACACTGCAAAGCATTTTTATGGACAAC GACCACAGGAATCTGGAATTTTATTAACATGGTCATCACTCCCCAAAATGGACGTTA CCAAATAGATTCCGATGTTCTCCTGATCCCCTGGAAGCTGACTTACAGGAATATTGGT TCTGATTTTATTCCTCGGGGCGCCTTTGGAAAGGTATACCTTGGCACAAGATATAAAGA CGAAGAAAAGAAATGGCGTGTAACATGATCCAGTAGATCAATTTAAGCCATCTGATGT GGAATCCAGGCTTGCTTCCGGCAGAGAACATCGCAGAGCTGTATGGCGCAGTCTCTG TGGGGTGAAACTGTCCATCTCTTTATGGAAGCAGGCGAGGGAGGGTCTGTTCTGGAGA AACTGGAGAGCTGTGACCAATGAGAGAATTTGAAATTATTGGGTGACAAAGCATCT		

	TCTCAAGGGACTTGATTTTCTACACTCAAAGAAAGTGATCCATCATGATATTAAACCT AGCAACATTGTTTTTCATGTCCACAAAAGCTGTTTTGGTGGATTTTGGCCTAAGTGTTT AAATGACCGAAGATGTCTATTTTCTAAGGACCTCCGAGGAACAGAGATTTACATGAG CCCAGAGGTCATCCTGTGCAGGGGCCATTCAACCAAAGCAGACATCTACAGCCTGGGG GCCACGCTCATCCACATGCAGACGGGCACCCACCCTGGGTGAAGCGCTACCTCGCT CAGCCTATCCCTCCTACCTGTACATAATCCACAAGCAAGCACCTCCACTGGAAGACAT TGCAGATGACTGCAGTCCAGGGATGAGAGAGCTGATAGAAGCTTCCTGGAGAGAAAC CCCAATCACCGCCCAAGAGCCGAGACCTACTAAAACATGAGGCCCTGAACCCGCCCA GAGAGGATCAGCCACGCTGTGAGAGTCTGGACTCTGCCCTCTTGAGCGCAAGAGGCT GCTGAGTAGGAAGGAGCTGGAACCTCCTGAGAACATTGCTTGAGCGGCCG		
	ORF Start: at 48		ORF Stop: TGA at 1317
	SEQ ID NO: 30	423 aa	MW at 48084.5kD
NOV1o, CG101683-05 Protein Sequence	EYMTGSDNKEEIDLLIKHLNVSDVIDIMENLYASEEPVYEPSLMTMCQDSNQNDER SKSLLLSGQVEPWLSSVRYGTVEDLLAFANHISNTAKHFYQRPQESGILLNMVITPQ NGRYQIDSDVLLIPWKLYRNIGSDFIPRGAFGKVYLAQDIKTKRMACKLIPVDQFK PSDVEIQACFRHENIAELYGAVLWGETVHLFMEAGEGGSVLEKLESCGPMREFEIIWV TKHVLKGLDFLHKKKVIHHDIKPSNIVFMSTKAVLVDFGLSVQMTEDVYFPKDLRGTE IYMSPEVILCRGHSTKADIYSLGATLIHMQTGTPPWVKRYPRSAIPSYLYIIHKQAPP LEDIADDCSPGMRELIEASLERNPNHRPRAADLLKHEALNPREDQPRCQSLDSALLE RKRLLSRKELELPENIA		
	SEQ ID NO: 31	1428 bp	
NOV1p, CG101683-06 DNA Sequence	ACCATGGGACATCATCACCACCATCAGGAGTACATGAGCACTGGAAGTGACAATAAG AAGAGATTGATTTATTAATTAACATTAAATGTGTCTGATGTAATAGACATTATGGA AAATCTTTATGCAAGTGAAGAGCCAGCAGTTTATGAACCCAGTCTAATGACCATGTGT CAAGACAGTAATCAAAACGATGAGCGTTCTAAGTCTCTGCTGCTTAGTGGCCAAGAGG TACCATGGTTGTCTCATCAGTACGATACGGAACGTGGAGGATTGTCTTGCTTTTGCAAA CCATATATCCAACACTGCAAAGCATTTTATGGACAACGACCACAGGAATCTGGAATT TTATTAACATGGTCATCACTCCCCAAAATGGACGTTACCAAATAGATTCCGATGTTT TCCTGATCCCTGGAAGCTGACTTACAGGAATATTGGTTCTGATTTTATTCCTCGGGG CGCCTTTGGAAAGGTATACTTGGCACAAGATATAAGACGAAGAAAAGAAATGGCGTGT AAACTGATCCCAGTAGATCAATTTAAGCCATCTGATGTGGAATCCAGGCTTGCTTCC GGCAGGAGAACATCGCAGAGCTGTATGGCGCAGTCTGTGGGGTGAAACTGTCCATCT CTTTATGGAAGCAGGCGAGGGAGGGTCTGTTCTGGAGAACTGGAGAGCTGTGGACCA ATGAGAGAATTTGAAATTATTTGGGTGACAAAGCATGTTCTCAAGGAGCTTGATTTTC TACACTCAAAGAAAGTGATCCATCATGATATTAAACCTAGCAACATTGTTTTCATGTC CACAAAAGCTGTTTTTGGTGGATTTTGGCCTAAGTGTTCAAATGACCGAAGATGTCTAT TTTCCTAAGGACCTCCGAGGAACAGAGATTTACATGAGCCAGAGGTCATCCTGTGCA GGGGCCATTCAACCAAAGCAGACATCTACAGCCTGGGGGCCACGCTCATCCACATGCA GACGGGCACCCACCTGGGTGAAGCGCTACCTCGCTCAGCCTATCCCTCTACCTG TACATAATCCACAAGCAAGCACCTCCACTGGAAGACATTGCAGATGACTGCAGTCCAG GGATGAGAGAGCTGATAGAAGCTTCCCTGGAGAGAAACCCCAATCACCGCCCAAGAGC CGCAGACCTACTAAAACATGAGGCCCTGAACCCGCCAGAGAGGATCAGCCACGCTGT CAGAGTCTGGACTCTGCCCTCTTGAGAGCGCAAGAGGCTGCTGAGTAGGAAGGAGCTGG AACTTCCTGAGAACATTGCTGATTCTTCGTGCACAGGAAGCACCGAGGAATCTGAGAT GCTCAAGAGGCAACGCTCTCTCTACATCGACCTCGGCGCTCTGGCTGGCTACTTCAAT CTTGTTCCGGGACCACCAACGCTTGAATATGGCTGA		
	ORF Start: at 1		ORF Stop: TGA at 1426
	SEQ ID NO: 32	475 aa	MW at 53904.9kD
NOV1p, CG101683-06 Protein Sequence	TMGHHHHHHEYMSTGSDNKEEIDLLIKHLNVSDVIDIMENLYASEEPVYEPSLMTMC QDSNQNDERSKSLLLSGQVEPWLSSVRYGTVEDLLAFANHISNTAKHFYQRPQESGI LLNMVITPQNGRYQIDSDVLLIPWKLYRNIGSDFIPRGAFGKVYLAQDIKTKRMAC KLIPVDQFKPSDVEIQACFRHENIAELYGAVLWGETVHLFMEAGEGGSVLEKLESCGP MREFEIIWVTKHVLKGLDFLHKKKVIHHDIKPSNIVFMSTKAVLVDFGLSVQMTEDVY FPKDLRGTEIYMSPEVILCRGHSTKADIYSLGATLIHMQTGTPPWVKRYPRSAIPSYL YIIHKQAPPLEDIADDCSPGMRELIEASLERNPNHRPRAADLLKHEALNPREDQPRC QSLDSALLERKRLLSRKELELPENIADSSCTGSTEESEMLKRQRSYIDLGALAGYFN LVRGPPTLEYG		

	SEQ ID NO: 33	1293 bp	
NOV1q, CG101683-07 DNA Sequence	GGGCCCCCTGGGATCCACCATGGAGTACATGAGCACTGGAAGTGACAATAAAGAAGAGA TTGATTATTAAATTAAACATTTAAATGTGTCTGATGTAATAGACATTATGGAAAATCT TTATGCAAGTGAAGAGCCAGCAGTTTATGAACCCAGTCTAATGACCATGTGTCAAGAC AGTAATCAAAACGATGAGCGTTCTAAGTCTCTGCTGCTTAGTGGCCAAGAGGTACCAT GGTTGTCATCAGTCAGATACGGAACGTGTGGAGGATTGCTTGTCTTTGCAAACCATAT ATCCAACACTGCAAAGCATTTTTATGGACAACGACCACAGGAATCTGGAATTTTATTA AACATGGTCATCACTCCCCAAAATGGACGTTACCAAATAGATTCCGATGTTCTCCTGA TCCCCTGGAAGCTGACTTACAGGAATATTGGTTCTGATTTTATTCTCGGGGCGCCTT TGGAAAGGTATACTTGGCACAAGATATAAAGACGAAGAAAAGAATGGCGTGAAACTG ATCCCAGTAGATCAATTTAAGCCATCTGATGTGGAAATCCAGGCTTGTCTCCGGCAG AGAACATCGCAGAGCTGTATGGCGCAGTCTGTGGGGTGAACTGTCCATCTCTTTAT GGAAGCAGGCGAGGGAGGGTCTGTTCTGGAGAACTGGAGAGCTGTGGACCAATGAGA GAATTTGAAATTTATTTGGGTGACAAAGCATGTTCTCAAGGGACTTGATTTTCTACACT CAAAGAAAGTGATCCATCATGATATTAACCTAGCAACATTGTTTTCTATGTCCACAAA AGCTGTTTTGGTGGATTTTGGCCTAAGTGTTCAAATGACCGAAGATGTCTATTTTCTCT AAGGACCTCCGAGGAACAGAGATTTACATGAGCCCAGAGGTATCTCTGTGCAGGGGCC ATTCAACCAAAGCAGACATCTACAGCCTGGGGGCCACGCTCATCCACATGCAGACGGG CACCACCCTGGGTGAAGCGCTACCTCGCTCAGCCTATCCCTCCTACCTGTACATA ATCCACAAGCAAGCACCTCCACTGGAAGACATTGCAGATGACTGCAGTCCAGGGATGA GAGAGCTGATAGAAGCTTCCCTGGAGAGAAACCCCAATCACCGCCAAGAGCCGAGAG CCTACTAAAACATGAGGCCCTGAACCCGCCCAGAGAGGATCAGCCACGCTGTACAGT CTGGACTCTGCCCTCTTGGAGCGCAAGAGGCTGCTGAGTAGGAAGGAGCTGGAACCTTC CTGAGAACATTGCTTGA		
	ORF Start: ATG at 19		ORF Stop: TGA at 1291
	SEQ ID NO: 34	424 aa	MW. at 48215.7kD
NOV1q, CG101683-07 Protein Sequence	MEYMTGSDNKEEIDLLIKHLNVSDVIDIMENLYASEEPAVYEPSLMTMCQDSNQND RSKSLLLSGQEV PWLSSVRYGTVEDLLAFANHISNTAKHFGYQRPQESGILLNMVITP QNGRYQIDSDVLLIPWKLYRNI GSDFI PRGAFGKVYLAQDIKTKRMAKCLI PVDQF KPSDVEIQACFRHENIAELYGAVLWGETVHLFMEAGEGGSVLEKLESCGPMREFEIIW VTKHVLKGLDFLHSHKVIHHDIKPSNIVFMSTKAVLVDFGLSVQMTEDVYFPKDLRGT EIYMSPEVILCRGHSTKADIYSLGATLIHMQTGTPPWVKRYPRSAVPSYLYIIHKQAP PLEDIADDCSPGMRELIEASLERNPNHRPRAADLLKHEALNPREDQPRCQSLDSALL ERKRLLSRKELELPENIA		
	SEQ ID NO: 35	1428 bp	
NOV1r, CG101683-08 DNA Sequence	CACCGCGCGCCGACCATGGAGTACATGAGCACTGGAAGTGACAATAAAGAAGAGATTG ATTTATTAATTAACATTTAAATGTGTCTGATGTAATAGACATTATGGAAAATCTTTA TGCAAGTGAAGAGCCAGCAGTTTATGAACCCAGTCTAATGACCATGTGTCAAGACAGT AATCAAAACGATGAGCGTTCTAAGTCTCTGCTGCTTAGTGGCCAAGAGGTACCATGGT TGTCATCAGTCAGATACGGAACGTGTGGAGGATTGCTTGTCTTTGCAAACCATATATC CAACACTGCAAAGCATTTTTATGGACAACGACCACAGGAATCTGGAATTTTATTAAC ATGGTCATCACTCCCCAAAATGGACGTTACCAAATAGATTCCGATGTTCTCCTGATCC CCTGGAAGCTGACTTACAGGAATATTGGTTCTGATTTTATTCTCGGGGCGCCTTTGG AAAGGTATACTTGGCACAAGATATAAAGACGAAGAAAAGAATGGCGTGTAACCTGATC CCAGTAGATCAATTTAAGCCATCTGATGTGGAAATCCAGGCTTGTCTCCGGCAGCAGA ACATCGCAGAGCTGTATGGCGCAGTCTGTGGGGTGAACTGTCCATCTCTTTATGGA AGCAGGCGAGGGAGGGTCTGTTCTGGAGAACTGGAGAGCTGTGGACCAATGAGAGAA TTTGAAATTATTTGGGTGACAAAGCATGTTCTCAAGGGACTTGATTTTCTACACTCAA AGAAAGTGATCCATCATGATATTAACCTAGCAACATTGTTTTCTATGTCCACAAAAGC TGTTTGGTGGATTTTGGCCTAAGTGTTCAAATGACCGAAGATGTCTATTTTCTAAG GACCTCCGAGGAACAGAGATTTACATGAGCCCAGAGGTATCTCTGTGCAGGGGCCATT CAACCAAAGCAGACATCTACAGCCTGGGGGCCACGCTCATCCACATGCAGACGGGCAC CCCACCCTGGGTGAAGCGCTACCTCGCTCAGCCTATCCCTCCTGATACATAATC CACAAGCAAGCACCTCCACTGGAAGACATTGCAGATGACTGCAGTCCAGGGATGAGAG AGCTGATAGAAGCTTCCCTGGAGAGAAACCCCAATCACCGCCAAGAGCCGACAGCCT ACTAAAACATGAGGCCCTGAACCCGCCCAGAGAGGATCAGCCACGCTGTACAGTCTG GACTCTGCCCTCTTGGAGCGCAAGAGGCTGCTGAGTAGGAAGGAGCTGGAACCTCCTG AGAACATTGCTGATTCTCTGTGCACAGGAAGCACCAGGAATCTGAGATGCTCAAGAG		

	GCAACGCTCTCTCTACATCGACCTCGGCGCTCTGGCTGGCTACTTCAATCTGTTCGG GGACCACCAACGCTTGAATATGGCTAGGTCGACGGC		
	ORF Start: ATG at 16		ORF Stop: TAG at 1417
	SEQ ID NO: 36	467 aa	MW at 52923.9kD
NOV1r, CG101683-08 Protein Sequence	MEYMSTGSDNKEEIDLLIKHLNVSDVIDIMENLYASEEPAVYEPSTMTMCQDSNQND RSKSLLLSGQEVPLSSVRYGTVEDLLAFANHISNTAKHFGQRPQESGILLNMVITP QNGRYQIDSDVLLIPWKLTYRNIGSDFIPRGAFGKVYLAQDIKTKRMACKLIIPVDQF KPSDVEIQACFRHENIAELYGAVLWGETVHLFMEAGEGGSVLEKLESCGPMREFEIIW VTKHVLKGLDFLHSHKVIHHDIKPSNIVFMSTKAVLVDFGLSVQMTEDVYFPKDLRGT EIYMSPEVILCRGHSTKADIYSLGATLIHMQTGTPPWVKRYPRSAVPSYLYIIHKQAP PLEDIADDCSPGMRELIEASLERNPNHRPRAADLLKHEALNPPREDQPCQSLDSALL ERKRLLSRKELELPENIADSSCTGSTEESEMLKRQRSYIDLGALAGYFNLVRGPPTL EYG		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 1B.

Table 1B. Comparison of NOV1a against NOV1b through NOV1r.		
Protein Sequence	NOV1a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV1b	1..467 2..468	466/467 (99%) 466/467 (99%)
NOV1c	1..424 5..428	423/424 (99%) 423/424 (99%)
NOV1d	1..467 2..468	466/467 (99%) 466/467 (99%)
NOV1e	2..467 13..478	465/466 (99%) 465/466 (99%)
NOV1f	1..424 2..425	423/424 (99%) 423/424 (99%)
NOV1g	2..424 16..438	422/423 (99%) 422/423 (99%)
NOV1h	2..467 10..475	465/466 (99%) 465/466 (99%)
NOV1i	1..467 5..471	466/467 (99%) 466/467 (99%)
NOV1j	1..467 20..486	466/467 (99%) 466/467 (99%)
NOV1k	1..467 20..486	466/467 (99%) 466/467 (99%)
NOV1l	1..467 1..467	466/467 (99%) 466/467 (99%)
NOV1m	1..467	466/467 (99%)

	2..468	466/467 (99%)
NOV1n	1..424 2..425	423/424 (99%) 423/424 (99%)
NOV1o	2..424 1..423	422/423 (99%) 422/423 (99%)
NOV1p	2..467 10..475	465/466 (99%) 465/466 (99%)
NOV1q	1..424 1..424	423/424 (99%) 423/424 (99%)
NOV1r	1..467 1..467	466/467 (99%) 466/467 (99%)

Further analysis of the NOV1a protein yielded the following properties shown in Table 1C.

Table 1C. Protein Sequence Properties NOV1a	
PSort analysis:	0.6500 probability located in cytoplasm; 0.1000 probability located in mitochondrial matrix space; 0.1000 probability located in lysosome (lumen); 0.0000 probability located in endoplasmic reticulum (membrane)
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV1a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 1D.

Table 1D. Geneseq Results for NOV1a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV1a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAE05951	Human cot oncoprotein encoded by D14497. oncogene - Homo sapiens, 467 aa. [US6265216-B1, 24-JUL-2001]	1..467 1..467	467/467 (100%) 467/467 (100%)	0.0
AAY79244	Human COT - Homo sapiens, 467 aa. [WO200011191-A2, 02-MAR-2000]	1..467 1..467	467/467 (100%) 467/467 (100%)	0.0
AAE10313	Human Tp12 protein - Homo sapiens, 467 aa. [WO200166559-A1, 13-SEP-2001]	1..467 1..467	466/467 (99%) 466/467 (99%)	0.0
AAE10314	Rat Tn12 protein - Rattus sp. 467	1..467	439/467 (94%)	0.0

	aa. [WO200166559-A1, 13-SEP-2001]	1..467	454/467 (97%)	
AAY79243	Rat TPL-2 - Rattus norvegicus, 467 aa. [WO200011191-A2, 02-MAR-2000]	1..467 1..467	438/467 (93%) 453/467 (96%)	0.0

In a BLAST search of public sequence databases, the NOV1a protein was found to have homology to the proteins shown in the BLASTP data in Table 1E.

Table 1E. Public BLASTP Results for NOV1a				
Protein Accession Number	Protein/Organism/Length	NOV1a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
P41279	Mitogen-activated protein kinase kinase kinase 8 (EC 2.7.1.-) (COT proto-oncogene serine/threonine-protein kinase) (C-COT) (Cancer Osaka thyroid oncogene) - Homo sapiens (Human), 467 aa.	1..467 1..467	467/467 (100%) 467/467 (100%)	0.0
A48713	serine/threonine-specific protein kinase cot, 58K form - human, 467 aa.	1..467 1..467	466/467 (99%) 466/467 (99%)	0.0
Q63562	Mitogen-activated protein kinase kinase kinase 8 (EC 2.7.1.-) (Tumor progression locus 2) (TPL-2) - Rattus norvegicus (Rat), 467 aa.	1..467 1..467	438/467 (93%) 453/467 (96%)	0.0
Q07174	Mitogen-activated protein kinase kinase kinase 8 (EC 2.7.1.-) (COT proto-oncogene serine/threonine-protein kinase) (C-COT) (Cancer Osaka thyroid oncogene) - Mus musculus (Mouse), 467 aa.	1..467 1..467	435/467 (93%) 454/467 (97%)	0.0
A41253	kinase-related transforming protein (EC 2.7.1.-) - human, 415 aa.	1..397 1..397	379/397 (95%) 379/397 (95%)	0.0

Pfam analysis predicts that the NOV1a protein contains the domains shown in the Table 1F.

Table 1F. Domain Analysis of NOV1a			
Pfam Domain	NOV1a Match Region	Identities/ Similarities for the Matched Region	Expect Value

pkinase	146..388	74/279 (27%) 187/279 (67%)	4.7e-54
---------	----------	-------------------------------	---------

**Example 2.**

The NOV2 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 2A.

Table 2A. NOV2 Sequence Analysis			
	SEQ ID NO: 37	917 bp	
NOV2a, CG101996-01 DNA Sequence	GATGAAGAGAGGGGAGCTCTTTGACTACCTCACTGAGAAGGTCACCTTGAGTGAGAAG GAAACCAGAAAGATCATGCGAGCTCTGCTGGAGGTGATCTGCACCTTGCACAACTCA ACATCGTGCACCGGACCTGAAGCCCCGAGAACATTCTCTTGGATGACAACATGAACAT CAAGCTCACAGACTTTGGCTTTTCTGCTGCCAGCTGGAGCCGGGAGAGAGGCTGCCGAGGG TCTGCGGGACCCCCAGTTACCTGGCCCCCTGAGATTATCGAGTGCTCCATGAATGAGGA CCACCCGGGCTACGGGAAGAGGTGGACATGTGGAGCACTGGCGTCAATCATGTACACG CTGCTGGCCGCTCCCCGCCCTTCTGGCACCGGAAGCAGATGCTGATGCTGAGGATGAT CATGAGCGGCAACTACAGTTTGGCTCGCCCCGAGTGGGATGATTACTCGGACACCGTG AAGGACCTGGTCTCCCGATTCTGGTGGTGCAACCCAGAACCGCTACACAGCGGAAG AGGCCTTGGCACACCCCTTCTTCCAGCAGTACTTGGTGGAGGAAGTGCGGCACCTTCAG CCCCGGGGGAAGTTCAAGGTGATCGCTCTGACCGTGCTGGCTTCAGTGCGGATCTAC TACCACTACCGCCGGGTGAAGCCTGTGACCCGGGAGATCGTCATCCGAGACCCCTATG CCCTCCGGCCTCTGCGCCGGCTCATCGACGCCTACGCTTTCCGAATCTATGGCCACTG GGTGAAGAAGGGGCAGCAGCAGAACCGGGCAGCCCTTTTCGAGAACACACCCAAGGCC GTGCTCCTCTCCCTGGCCGAGGAGGACTACTGAGGGGCTGGCCAGTCAGGGAGGGCTA GGGGCAGGTGGGAGGGGAAGCCATGGAATACAAGTCAAAGGGGT		
	ORF Start: ATG at 387		ORF Stop: TGA at 843
	SEQ ID NO: 38	152 aa	MW at 18023.7kD
NOV2a, CG101996-01 Protein Sequence	MLMLRMIMSGNYQFGSPEDDYSDTVKDLVSRFLVVQPQNRYTAEELAHPPFQQYLV EEVRHFSPRGKFKVIALTVLASVRIYYQYRRVKPVTREIVIRDPYALRPLRLRIDAYA FRIYGHVWKKGQQQNRAALFENTPKAVLLSLAEEDY		
	SEQ ID NO: 39	1299 bp	
NOV2b, CG101996-04 DNA Sequence	ATGACCCGGGACGAGGCACTGCCGACTCTCATTCTGCACAGGACTTCTATGAGAATT ATGAGCCCAAAGAGATCCTGGGCAGGGGCGTTAGCAGTGTGGTCAGGCGATGCATCCA CAAGCCACGAGCCAGGAGTACGCCGTGAAGGTATCGACGTCACCGGTGGAGGCGAGC TTCAGCCCGGAGGAGGTGCGGGAGCTGCGAGAAGCCACGCTGAAGGAGGTGGACATCC TGCGCAAGGTCTCAGGGCACCCCAACATCATAAGCTGAAGGACACTTATGAGACCAA CACTTTCTTCTTCTTGGTGTGTTGACCTGATGAAGAGAGGGGAGCTCTTTGACTACCTC ACTGAGAAGGTCACCTTGAGTGAGAAGGAAACCAGAAAGATCATGCGAGCTCTGCTGG AGGTGATCTGCACCTTGACAAACTCAACATCGTGACCCGGGACCTGAAGCCCCGAGAA CATTCTCTTGGATGACAACATGAACATCAAGCTCACAGACTTTGGCTTTTCTGCCAG CTGGAGCCGGGAGAGAGGCTGCGAGTAGAGACAGGGTTTACCATGTTGGTCAGGCTG GTCTCGAATCCTGACCTTACGATCCGCCCGCTCGGCCTCCCAAAGTGCTGTGATTA CAGGCGTGAGCCACCATGCCCAGCAGGGCTAGGCATTTCTTACAGGCTGCGGGAGCC CCCAGTTACCTGGCCCCCTGAGATTATCGAGTGCTCCATGAATGAGGACCACCCGGGCT ACGGGAAAGAGGTGGACATGTGGAGCACTGGCGTCAATCATGTACACGCTGCTGGCCGG CTCCCCGCCCTTCTGGCACCGGAAGCAGATGCTGATGCTGAGGATGATCATGAGCGGC AACTACCAGTTTGGCTCGCCCGAGTGGGATGATTACTCGGACACCGTGAAGGACCTGG TCTCCCGATTCTGGTGGTGCAACCCAGAACCGCTATACAGCGGAAGAGGCCCTTGGC ACACCCCTTCTCCAGCAGTACTTGGTAGAGGAAGTGCGGCACCTTACGCCCCGGGGG AAGTTCAAGGTGATCGCTCTGACCGTGCTGGCTTCAGTGCGGATCTACTACCAGTACC GCCGGGTGAAGCCTGTGACCCGGGAGATCGTCATCCGAGACCCCTATGCCCTCCGGCC TCTGCGCCGGCTCATCGACGCCTACGCTTTCCGAATCTATGGCCACTGGGTGAAGAAG GGGCAGCAGCAGAACCGGGCAGCCCTTTTCGAGAACACACCCAAGGCCGTGCTCCTCT		

	CCCTGGCCGAGGAGGACTACTGA		
	ORF Start: ATG at 1		ORF Stop: TGA at 1297
	SEQ ID NO: 40	432 aa	MW at 49811.7kD
NOV2b, CG101996-04 Protein Sequence	MTRDEALPDSHSAQDFYENYEPKEILGRGVSSVVRRCIHKPTSQEYAVKVIDVTGGGS FSPEEVRELREATLKEVDILRKVSGHPNIIQLKDTYETNTFFFLVFDLMKRGELFDYL TEKVTLSSEKETRKIMRALLEVICTLHKLNI VHRDLKPENILLDDNMNIKLTDGFGSCQ LEPGERLRVETGFHHVQGAGLELLTLRSARLGLPKCCDYRREPPCPAGLGISSEVCGT PSYLAPEIIIECSMNEDHPGYGKEVDMWSTGVIMYTLLAGSPFFWHRKQMLMLRMIMSG NYQFGSPEWDDYSDTVKDLVSRFLVVQPQNRYTAEALAHPPFQQYLVEVRHFSRPG KFKVIALTVLASVRIYYQYRRVKPVTREIVIRDPYALRPLRLIDAYAFRIYGHVWKK GQQQNRAALFENTPKAVLLSLAEEDY		
	SEQ ID NO: 41	1377 bp	
NOV2c, CG101996-02 DNA Sequence	GGCCTTCAGCCCTCTGTGGTCCCCTCTCCCGGGGGGCTTTGGGATCTTGTCAAGCT CCTTCAAGAGCCTGCAAGCACTTAACCAGCCACCCAGAGTTCCTCACTGAAGATCTG AGCATGACCCGGGACGAGGCACTGCCGACTCTCATTCTGCACAGGACTTCTATGAGA ATTATGAGCCCAAAGAGATCCTGGGCAGGGGCGTTAGCAGTG7GGTCAGGCGATGCAT CCACAAGCCACGAGCCAGAGTACGCCGTGAAGGTCATCGACGTCACCGTGGAGGC AGCTTCAGCCCGGAGGAGGTGCGGGAGCTGCGAGAAGCCACGCTGAAGGAGGTGACA TCCTGCGCAAGGTCTCAGGCAACCCCAACATCATAACAGCTGAAGGACACTTATGAGAC CAACACTTTCTTCTTCTTGTGTTTGACCTGATGAAGAGAGGGGAGCTCTTTGACTAC CTCACTGAGAAGGTACCTTGAGTGAGAAGGAAACAGAAAGATCATGCGAGCTCTGC TGGAGGTGATCTGCACCTTGCACAACTCAACATCGTGCACCGGACCTGAAGCCCGA GAACATTCTCTTGGATGACAACATGAACATCAAGCTCACAGACTTGGCTTTTCTTGC CAGCTGGAGCCGGGAGAGAGGCTGCGAGAGGTCTGCGGACCCCAAGTTACCTGGCCC CTGAGATTATCGAGTGCTCCATGAATGAGGACCACCCGGCTACGGGAAAGAGGTGGA CATGTGGAGCACTGGCGTCATCATGTACACGCTGCTGGCCGGCTCCCGCCCTTCTTG CACCAGGAGCAGATGCTGATGCTGAGGATGATCATGAGCGGCAACTACCAAGTTTGGCT CGCCCGAGTGGGATGATTACTCGGACACCGTGAAGGACCTGGTCTCCCGATTCTTGGT GGTGCAACCCAGAACCGCTACACAGCGGAAGAGGCTTGGCACACCCCTTCTTCCAG CAGTACTTGGTGGAGGAAGTGCAGGCACTTCAGCCCCGGGGGAAGTTCAAGGTGATCG CTCTGACCGTGTGGCTTCACTGCGGATCTACTACCAAGTACCGCCGGGTGAAGCCTGT GACCCGGGAGATCGTCATCCGAGACCCCTATGCCCTCCGGCTCTGCGCCGGCTCATC GACGCTACGCTTTTCCGAATCTATGGCCACTGGGTGAAGAAGGGGAGCAGCAGAAACC GGGAGCCCTTTTCGAGAACACACCAAGGCCGTGCTCCTCTCCCTGGCCGAGGAGGA CTACTGAGGGGCTGGCCAGTCAGGAGGGCTAGGGGGCAGGTGGGGAGGGGAAGCCAT GGAAATACAAGTCAAAGGGGTAAAAAAAAAAAAAAAAAAAAA		
	ORF Start: ATG at 120		ORF Stop: TGA at 1281
	SEQ ID NO: 42	387 aa	MW at 45023.3kD
NOV2c, CG101996-02 Protein Sequence	MTRDEALPDSHSAQDFYENYEPKEILGRGVSSVVRRCIHKPTSQEYAVKVIDVTGGGS FSPEEVRELREATLKEVDILRKVSGHPNIIQLKDTYETNTFFFLVFDLMKRGELFDYL TEKVTLSSEKETRKIMRALLEVICTLHKLNI VHRDLKPENILLDDNMNIKLTDGFGSCQ LEPGERLRVETGFHHVQGAGLELLTLRSARLGLPKCCDYRREPPCPAGLGISSEVCGT PSYLAPEIIIECSMNEDHPGYGKEVDMWSTGVIMYTLLAGSPFFWHRKQMLMLRMIMSG NYQFGSPEWDDYSDTVKDLVSRFLVVQPQNRYTAEALAHPPFQQYLVEVRHFSRPG KFKVIALTVLASVRIYYQYRRVKPVTREIVIRDPYALRPLRLIDAYAFRIYGHVWKK GQQQNRAALFENTPKAVLLSLAEEDY		
	SEQ ID NO: 43	1165 bp	
NOV2d, 245245680 DNA Sequence	CATGACCCGGGACGAGGCACTGCCGACTCTCATTCTGCACAGGACTTCTATGAGAAT TATGAGCCCAAAGAGATCCTGGGACGGGCGTTAGCAGTGTGGTCAGGCGATGCATCC ACAAGCCACGAGCCAGGAGTACGCCGTGAAGGTATCGACGTCACCGTGGAGGAG CTTCAGCCCGGAGGAGGTGCGGGAGCTGCGAGAAGCCACGCTGAAGGAGGTGGACATC CTGCGCAAGGTCTCAGGGCACCCCAACATCATAACAGCTGAAGGACACTTATGAGACCA ACACTTTCTTCTTCTTGGTGTGTTGACCTGATGAAGAGAGGGGAGCTCTTTGACTACCT CACTGAGAAGGTACCTTGAGTGAGAAGGAAACAGAAAGATCATGCGAGCTCTGCTG GAGGTGATCTGCACCTTGCACAACTCAACATCGTGCACCGGACCTGAAGCCCGAGA ACATTCTCTTGGATGACAACATGAACATCAAGCTCACAGACTTTGGCTTTTCTTCCCA GCTGGAGCCGGGAGAGAGGCTGCGAGAGGTCTGCGGGACCCCAAGTTACCTGGCCCT		

	GAGATTATCGAGTGCTCCATGAATGAGGACCACCCGGGCTACGGGAAAGAGGTGGACA TGTGGAGCACTGGCGTCATCATGTACACGCTGCTGGCCGGCTCCCCGCCCTTCTGGCA CCGGAAGCAGATGCTGATGCTGAGGATGATCATGAGCGGCAACTACCAGTTTGGCTCG CCCGAGTGGGATGATTACTCGGACACCGTGAAGGACCTGGTCTCCCGATTCTGGTGG TGCAACCCAGAACCGCTACACAGCGGAAGAGGCCTTGGCACACCCCTTCTTCCAGCA GTACTTGGTGGAGGAAGTGC GGCACTTCAGCCCCCGGGGGAAGTTCAAGGTGATCGCT CTGACCGTGCTGGCTTCAGTGGGATCTACTACCAGTACCGCCGGGTGAAGCCTGTGA CCCGGGAGATCGTCATCCGAGACCCCTATGCCCTCCGGCCTCTGCGCCGGCTCATCGA CGCCTACGCTTTCGAATCTATGGCCACTGGGTGAAGAAGGGGCAGCAGCAGAACCGG GCAGCCCTTTTCGAGAACACACCCAAGGCCGTGCTCTCTCCCTGGCCGAGGAGGACT ACTGA		
	ORF Start: ATG at 2		ORF Stop: TGA at 1163
	SEQ ID NO: 44	387 aa	MW at 45023.3kD
NOV2d, 245245680 Protein Sequence	MTRDEALPDSHAQDFYENYEPKEILGRGVSSVVRRCIHKPTSQEYAVKVIDVTGGGS FSPEEVRELREATLKEVDILRKVSGHPNIIQLKDYETNTFFFLVFDLMKRGELFDYL TEKVTLSEKETRKIMRALLEVICTLHKLNI VHRDLKPENILLDDNMNIKLTDGFGSCQ LEPGERLREVCCTPSYLAPEII ECSMNEDHPGYGKEVDMWSTGVIMYTLLAGSPFFWH RKQMLMLRMIMSGNYQFGSPWDDYSDTVKDLVSRFLVVPQPNRYTAEALAHPPFQQ YLVEEVRHFSRPGKFKVIALTVLASVRIYYQYRRVKPVTREIVIRDPYALRPLRLRID AYAFRIYGHVWKKGQQQNRAALFENTPKAVLLSLAEEDY		
	SEQ ID NO: 45	1300 bp	
NOV2e, 245245707 DNA Sequence	CATGACCCGGGACGAGGCACTGCCGACTCTCATTCTGCACAGGACTTCTATGAGAAT TATGAGCCCAAAGAGATCCTGGGCAGGGGCGTTAGCAGTGTGGTCAGGCGATGCATCC ACAAGCCACGAGCCAGGAGTACGCCGTGAAGGTCATCGACGTCACCGGTGGAGGCAG CTTCAGCCCGGAGGAGGTGCGGGAGCTGCGAGAAGCCACGCTGAAGGAGGTGGACATC CTGCGCAAGGTCTCAGGGCACCCCAACATCATACAGCTGAAGGACACTTATGAGACCA ACACTTTCTTCTTCTTGGTGTTTGACCTGATGAAGAGAGGGGAGCTCTTTGACTACCT CACTGAGAAGGTCACCTTGAGTGAGAAGGAACCAGAAAGATCATGCGAGCTCTGCTG GAGGTGATCTGCACCTTGCACAACTCAACATCGTGCACCGGGACCTGAAGCCCGAGA ACATTCTCTTGGATGACAACATGAACATCAAGCTCACAGACTTTGGCTTTTCTGCCA GCTGGAGCCGGGAGAGAGGCTGCGAGTAGAGACAGGGTTTACCATTGTGGTCAGGCT GGTCTCAAACCTCTGACCTTACGATCCGCCCGCCTCGGCCTCCCAAAGTGCTGTGATT ACAGGCGTGAGCCACCATGCCAGCAGGGCTAGGCATTCTCAGAGGTCTCGGGGAC CCCCAGTTACCTGGCCCTGAGATTATCGAGTGCTCCATGAATGAGGACCAACCGGCG TACGGGAAAGAGGTGGACATGTGGAGCACTGGCGTCATCATGTACACGCTGCTGGCCG GCTCCCCGCCCTTCTGGCACCGGAAGCAGATGCTGATGCTGAGGATGATCATGAGCGG CAACTACCAGTTTGGCTCGCCCGAGTGGGATGATTACTCGGACACCGTGAAGGACCTG GTCTCCCGATTCTTGGTGGTGCAACCCAGAACCGCTACACAGCGGAAGAGGCCCTTGG CACACCCCTTCTTCCAGCAGTACTTGGTGGAGGAAGTGGCGCACTCAGCCCCGGGG GAAGTTCAAGGTGATCGCTCTGACCGTGCTGGCTTCAGTGGGATCTACTACCAGTAC CGCCGGGTGAAGCCTGTGACCGGGAGATCGTCATCCGAGACCCCTATGCCCTCCGGC CTCTGCGCCGGCTCATCGACGCTACGCTTTCGAATCTATGGCCACTGGGTGAAGAA GGGGCAGCAGCAGAACCGGGCAGCCCTTTTCGAGAACACACCCAAGGCCGTGCTCCTC TCCCTGGCCGAGGAGGACTACTGA		
	ORF Start: ATG at 2		ORF Stop: TGA at 1298
	SEQ ID NO: 46	432 aa	MW at 49810.8kD
NOV2e, 245245707 Protein Sequence	MTRDEALPDSHAQDFYENYEPKEILGRGVSSVVRRCIHKPTSQEYAVKVIDVTGGGS FSPEEVRELREATLKEVDILRKVSGHPNIIQLKDYETNTFFFLVFDLMKRGELFDYL TEKVTLSEKETRKIMRALLEVICTLHKLNI VHRDLKPENILLDDNMNIKLTDGFGSCQ LEPGERLRVETGFHHVQAGLKLTLRSARLGLPKCCDYRREPPCPAGLGISSEVCGT PSYLAPEII ECSMNEDHPGYGKEVDMWSTGVIMYTLLAGSPFFWHRKQMLMLRMIMSG NYQFGSPWDDYSDTVKDLVSRFLVVPQPNRYTAEALAHPPFQQYLVEEVRHFSRPG KFKVIALTVLASVRIYYQYRRVKPVTREIVIRDPYALRPLRLRIDAYAFRIYGHVWKK GQQQNRAALFENTPKAVLLSLAEEDY		
	SEQ ID NO: 47	927 bp	
NOV2f,	ACCATGGGACATCATCACCACATCACACCCGGGACGAGGCACTGCCGACTCTCATT		

248494552 DNA Sequence	CTGCACAGGACTTCTATGAGAATTATGAGCCCAAAGAGATCCTGGGCAGGGGCGTTAG CAGTGTGGTCAGGCGATGCATCCACAAGCCCACGAGCCAGGAGTACGCCGTGAAGGTC ATCGACGTCACCGGTGGAGGCAGCTTCAGCCCGGAGGAGGTGCGGGAGCTGCGAGAAG CCACGCTGAAGGAGGTGGACATCCTGCGCAAGGTCTCAGGGCACCACATCATACA GCTGAAGGACACTTATGAGACCAACACTTCTTCTTCTTGGTGTGTGACCTGATGAAG AGAGGGGAGCTCTTTGACTACCTCACTGAGAAGGTCACCTTGAGTGAGAAGGAAACCA GAAAGATCATGCGAGCTCTGCTGGAGGTGATCTGCACCTTGCACAACTCAACATCGT GCACCGGGACCTGAAGCCCGAGAACATTCTTGGATGACAACATGAACATCAAGCTC ACAGACTTTGGCTTTTCTGCCAGCTGGAGCCGGGAGAGAGGCTGCGAGAGGTCTGCG GGACCCCCAGTTACCTGGCCCCTGAGATTATCGAGTGTCCATGAATGAGGACCACCC GGGCTACGGGAAAGAGGTGGACATGTGGAGCACTGGCGTCATCATGTACAGCTGCTG GCCGGCTCCCCGCCCTTCTGGCACCGGAAGCAGATGCTGATGCTGAGGATGATCATGA GCGGCACTACCAGTTTGGCTCGCCCCAGTGGGATGATTACTCGGACACCGTGAAGGA CCTGGTCTCCCGATTCTTGGTGGTGCAACCCAGAACCCTACACAGCGGAAGAGGCC TTGGCACACCCCTTCTTCCAGCAGTACTTGGTGGAGGAAGTGCGGCACTTCACTGA		
	ORF Start: at 1		ORF Stop: TGA at 925
	SEQ ID NO: 48	308 aa	MW at 35743.4kD
NOV2f, 248494552 Protein Sequence	TMGHHHHHTRDEALPDSHSAQDFYENYEPKEILGRGVSSVVRRCIHKPTSQEYAVKV IDVTGGGSFSPPEVRELREATLKEVDILRKVSGHPNIIQLKDTYETNTFFFLVFDLMK RGELFDYLTEKVTLSSEKETRKIMRALLEVICTLHLKNI VHRDLKPENILLDDNMNIKL TDFGFSCQLEPGERLREVCCTPSYLAPEII ECSMNEDHPGYGKEVDMWSTGVIMYTLL AGSPPFWHRKQMLMLRMIMSGNYQFGSPWDDYSDTVKDLVSRFLVQFPQNRYTAEBA LAHPFFQQYLVEEVRHFS		
	SEQ ID NO: 49	1194 bp	
NOV2g, 242435676 DNA Sequence	CGCGGATCCACCATGACCCGGGACGAGGCACTGCCGGACTCTCATTCTGCACAGGACT TCTATGAGAATTATGAGCCCAAAGAGATCCTGGGCAGGGGCGTTAGCAGTGTGGTCAG GCGATGCATCCACAAGCCCACGAGCCAGGAGTACGCCGTGAAGGTATCGACGTCACC GGTGGAGGCAGCTTCAGCCCGGAGGAGGTGCGGGAGCTGCGAGAAGCCACGCTGAAGG AGGTGGACATCCTGCGCAAGGTCTCAGGGCACCCCAACATCATACAGCTGAAGGACAC TTATGAGACCAACACTTTCTTCTTCTTGGTGTGTGACCTGATGAAGAGAGGGGAGCTC TTTGACTACCTCACTGAGAAGGTACCTTGAGTGAGAAGGAAACCAGAAAGATCATGC GAGCTCTGCTGGAGGTGATCTGCACCTTGCACAACTCAACATCGTGCACCGGGACCT GAAGCCCGAGAACATTCTCTTGGATGACAACATGAACATCAAGCTCAGACACTTTGGC TTTTCTGCGCAGCTGGAGCCGGGAGAGAGGCTGCGAGAGGTCTGCGGGACCCCCAGTT ACCTGGCCCCCTGAGATTATCGAGTGTCTCATGAATGAGGACCACCCGGGCTACGGGAA AGAGGTGGACATGTGGAGCACTGGCGTCATCATGTACACGCTGCTGGCCGGCTCCCCG CCCTTCTGGCACCGGAAGCAGATGCTGATGCTGAGGATGATCATGAGCGGCAACTACC AGTTTGGCTCGCCGAGTGGGATGATTACTCGGACACCGTGAAGGACCTGGTCTCCCG ATTCTGGTGGTGCAACCCAGAACCCTACACAGCGGAAGAGGCCCTTGGCACACCCC TTCTTCCAGCAGTACTTGGTGGAGGAAGTGCGGCACTTCAGCCCCCGGGGAAGTTCA AGGTGATCGCTCTGACCGTGTCTGGCTTCACTGCGGATCTACTACCAGTACCGCCGGT GAAGCCTGTGACCCGGGAGATCGTCATCCGAGACCCCTATGCCCTCCGGCCTCTGCGC CGGCTCATCGACGCTACGCTTTCGAATCTATGGCCACTGGGTGAAGAAGGGGACG AGCAGAACCGGGCAGCCCTTTTCGAGAACACACCCCAAGCCGTGCTCTCTCCCTGGC CGAGGAGGACTACTGAGCGGCCCTTTTTCCTT		
	ORF Start: at 1		ORF Stop: TGA at 1174
	SEQ ID NO: 50	391 aa	MW at 45424.7kD
NOV2g, 242435676 Protein Sequence	RGSTMTRDEALPDSHSAQDFYENYEPKEILGRGVSSVVRRCIHKPTSQEYAVKVIDVT GGGSFSPPEVRELREATLKEVDILRKVSGHPNIIQLKDTYETNTFFFLVFDLMKRGEL FDYLTEKVTLSSEKETRKIMRALLEVICTLHLKNI VHRDLKPENILLDDNMNIKLTDG FSCQLEPGERLREVCCTPSYLAPEII ECSMNEDHPGYGKEVDMWSTGVIMYTLLAGSP PFWHRKQMLMLRMIMSGNYQFGSPWDDYSDTVKDLVSRFLVQFPQNRYTAEALAHF FFQQYLVEEVRHFSRPGKFKVIALTVLASVRIYYQYRRVKPVTREIVIRDPYALRPLR RLIDAYAFRIYGHVKKGQQQNRAALFENTPKAVLLSLAEDY		
	SEQ ID NO: 51	952 bp	
NOV2h,	ACATCATCACCACCATCACACCCGGGACGAGGCACTGCCGGACTCTCATTCTGCACAG		

254868664 DNA Sequence	GACTTCTATGAGAATTATGAGCCCAAAGAGATCCTGGGCAGGGGCGTTAGCAGTGTGGTCAGGCGATGCATCCACAAGCCCACGAGCCAGGAGTACGCCGTGAAGGTTCATCGACGT CACCGGTGGAGGCAGCTTCAGCCCGGAGGAGGTGCGGGAGCTGCGAGAAGCCACGCTG AAGGAGGTGGACATCCTGCGCAAGGTCTCAGGGCACCCCAACATCATACAGCTGAAGG ACACCTTATGAGACCAACACTTTCTTCTTCTTGGTGTGTTGACCTGATGAAGAGAGGGGA GCTCTTTGACTACCTCACTGAGAAGGTACCTTGAGTGAGAAGGAAACCAGAAAGATC ATGCGAGCTCTGCTGGAGGTGATCTGCACCTTGCAAACTCAACATCGTGACCCGGG ACCTGAAGCCCGAGAACATTCTCTTGGATGACAACATGAACATCAAGCTCACAGACTT TGGCTTTTCTCTGCCAGCTGGAGCCGGGAGAGAGGCTGCGAGAGGTCTGCGGGACCCCC AGTTACCTGGCCCCCTGAGATTATCGAGTGCTCCATGAATGAGGACCACCCGGGCTACG GGAAGAGGTGGACATGTGGAGCACTGGCGTCATCATGTACACGCTGCTGGCCGGCTC CCCGCCCTTCTGGCACCGGAAGCAGATGCTGTGCTGAGGATGATCATGAGCGGCAAC TACCAGTTTGGCTCGCCGAGTGGGATGATTACTCGGACACCGTGAAGGACCTGGTCT CCCGATTCTTGGTGGTGCAACCCAGAACCGCTACACAGCGGAAGAGGCCTTGGGCACA CCCCTTCTTCAGCAGTACTTGGTGGAGGAAGTGCGGCACCTTCAGCTGAGCGGCCGCA CTCGAGCACCACCACCACCAC		
	ORF Start: at 2		ORF Stop: TGA at 917
	SEQ ID NO: 52	305 aa	MW at 35454.0kD
NOV2h, 254868664 Protein Sequence	HHHHHHRDEALPDHSAQDFYENYEPKEILGRGVSSVVRRCIHKPTSQEYAVKVIDV TGGGSFSPPEEVRELREATLKEVDILRKVSGHPNIIQLKDTYETNTFFFLVFDLMKRGE LFDYLTEKVTLSEKETRKIMRALLEVICTLHLKLNIVHRDLKPENILLDDNMNIKLTDF GFSCQLEPGERLREVCCTPSYLAPEIIECMNEDHPGYGKEVDMWSTGVIMYTLLAGS PPFWHRKQMLMLRMIMSGNYQFGSPWDDYSDTVKDLVSRFLVVQPQPNRYTAEALAH PFFQQLVEEVRHFS		
	SEQ ID NO: 53	939 bp	
NOV2i, 249122191 DNA Sequence	CATATGACCCGGGACGAGGCACTGCCGACTCTCATTCTGCACAGGACTTCTATGAGA ATTATGAGCCCAAAGAGATCCTGGGCAGGGGCGTTAGCAGTGTGGTCAGGCGATGCAT CCACAAGCCCACGAGCCAGGAGTACGCCGTGAAGGTTCATCGACGTCACCGGTGGAGGC AGCTTCAGCCCGGAGGAGGTGCGGGAGCTGCGAGAAGCCACGCTGAAGGAGGTGGACA TCCTGCGCAAGGTCTCAGGGCACCCCAACATCATACAGCTGAAGGACACTTATGAGAC CAACACTTTCTTCTTCTTGGTGTGTTGACCTGATGAAGAGAGGGGAGCTCTTTGACTAC CTCAC TGAGAAGGTACCTTGAGTGAGAAGGAAACCAGAAAGATCATGCGAGCTCTGC TGGAGGTGATCTGCACCTTGCAAACTCAACATCGTGACCCGGGACCTGAAGCCCGA GAACATTCTCTTGGATGACAACATGAACATCAAGCTCACAGACTTTGGCTTTTCTCTGC CAGCTGGAGCCGGGAGAGAGGCTGCGAGAGGTCTGCGGGACCCCCAGTTACCTGGCCC CTGAGATTATCGAGTGCTCCATGAATGAGGACCACCCGGGCTACGGGAAAGAGGTGGA CATGTGGAGCACTGGCGTCATCATGTACACGCTGCTGGCCGGCTCCCGCCCTTCTGG CACCGGAAGCAGATGCTGATGCTGAGGATGATCATGAGCGGCAACTACCAGTTTGGCT CGCCCGAGTGGGATGATTACTCGGACACCGTGAAGGACCTGGTCTCCCGATTCTTGGT GGTGCAACCCAGAACCGCTACACAGCGGAAGAGGCCTTGGCACACCCCTTCTTCCAG CAGTACTTGGTGGAGGAAGTGCGGCACCTTCAGCTGAGCGGCCGCACTCGAGCACCACC ACCACCAC		
	ORF Start: at 1		ORF Stop: TGA at 904
	SEQ ID NO: 54	301 aa	MW at 34899.5kD
NOV2i, 249122191 Protein Sequence	HMTRDEALPDHSAQDFYENYEPKEILGRGVSSVVRRCIHKPTSQEYAVKVIDVTGGG SFSPEEVRELREATLKEVDILRKVSGHPNIIQLKDTYETNTFFFLVFDLMKRGE LFDY LTEKVTLSEKETRKIMRALLEVICTLHLKLNIVHRDLKPENILLDDNMNIKLTDFGFSC QLEPGERLREVCCTPSYLAPEIIECMNEDHPGYGKEVDMWSTGVIMYTLLAGSPPFW HRRKQMLMLRMIMSGNYQFGSPWDDYSDTVKDLVSRFLVVQPQPNRYTAEALAH PFFQ QYLVEEVRHFS		
	SEQ ID NO: 55	951 bp	
NOV2j, 249122234 DNA Sequence	ACCCGGGACGAGGCACTGCCGACTCTCATTCTGCACAGGACTTCTATGAGAATTATG AGCCCAAAGAGATCCTGGGCAGGGGCGTTAGCAGTGTGGTCAGGCGATGCATCCACAA GCCACGAGCCAGGAGTACGCCGTGAAGGTTCATCGACGTCACCGGTGGAGGCAGCTTC AGCCCGGAGGAGGTGCGGGAGCTGCGAGAAGCCACGCTGAAGGAGGTGGACATCCTGC GCAAGGTCTCAGGGCACCCCAACATCATACAGCTGAAGGACACTTATGAGACCAACAC		

	TTTCTTCTTCTTGGTGTGACCTGATGAAGAGAGGGGAGCTCTTTGACTACCTCACT GAGAAGGTACCTTGTAGTGAGAAGGAAACCAGAAAGATCATGCGAGCTCTGCTGGAGG TGATCTGCACCTTGACAACTCAACATCGTGACCGGGACCTGAAGCCCCGAGAACAT TCTCTTGGATGACAACATGAACATCAAGCTCACAGACTTTGGCTTTTCTGCCAGCTG GAGCCGGGAGAGAGGCTGCGAGAGGTCTGCGGGACCCCCAGTTACCTGGCCCCTGAGA TTATCGAGTGCTCCATGAATGAGGACCACCGGGCTACGGGAAAGAGGTGGACATGTG GAGCACTGGCGTCATCATGTACACGCTGCTGGCCGGCTCCCCGCCCTTCTGGCACCGG AAGCAGATGCTGATGCTGAGGATGATCATGAGCGGCAACTACCAGTTTGGCTCGCCG AGTGGGATGATTACTCGACACCGTGAAGGACCTGGTCTCCCGATTCTGGTGGTGCA ACCCAGAACCGCTACACAGCGGAAGAGGCCTTGGCACACCCCTTCTTCCAGCAGTAC TTGGTGGAGGAAGTGGGCACCTCAGCCATCATCACCACCATCACTGAGCGGCCGCAC TCGAGCACCACCACCACCAC		
	ORF Start: at 1		ORF Stop: TGA at 916
	SEQ ID NO: 56	305 aa	MW at 35454.0kD
NOV2j, 249122234 Protein Sequence	TRDEALPDSHSAQDFYENYEPKEILGRGVSSVVRRCIHKPTSQYAVKVIDVTGGGSF SPPEEVRELREATLKEVDILRKVSGHPNIIQLKDTYETNTFFFLVFDLMKRGELFDYLT EKVTLSEKETRKIMRALLEVICTLHKLNIHVHDLKPENILLDDNMNIKLTDGFGSCQL EPGERLREVCCTPSYLAPEIIECMNEHDHPGYGKEVDMWSTGVIMYTLLAGSPPFWR KQMLMLRMIMSGNYQFGSPWDDYSDTVKDLVSRFLVVQPQNRITYAEALAHPPFQQY LVEEVVRFHSHHHHHH		
	SEQ ID NO: 57	1252 bp	
NOV2k, CG101996-03 DNA Sequence	CTTTGGGATCTTGTCAAGCTCCTTCAAGAGCCTGCAAGCATTAAACCAGCCACCCAG AGTTCCCTCACTGAAGATCTGAGCATGACCCGGGACGAGGCACTGCCGGACTCTCATT CTGCACAGGACTTCTATGAGAATTATGAGCCCAAAGAGATCTGGGCAGGGCGGTAG CAGTGTGGTCAGGCGATGCATCCACAAGCCCACGAGCCAGGAGTACGCCGTGAAGGTC ATCGACGTCACCGGTGGAGGCAGCTTCAGCCCGGAGGAGGTGCGGGAGCTGCGAGAAG CCACGCTGAAGGAGGTGGACATCTGCGCAAGGTCACAGGACCCCAACATCATACA GCTGAAGGACACTTATGAGACCAACACTTTCTTCTTCTTGGTGTGTTGACTGATGAAG AGAGGGGAGCTCTTTGACTACCTCACTGAGAAGGTACCTTGAGTGAGAAGGAAACCA GAAAGATCATGCGAGCTCTGCTGGAGGTGATCTGCACCTTGACAACTCAACATCGT GCACCGGGACCTGAAGCCCAGAACATTCTCTTGGATGACAACATGAACATCAAGCTC ACAGACTTTGGCTTTTCTGCGAGCTGGAGCCGGGAGAGAGGCTGCGAGAGGTCTGCG GGACCCCCAGTTACCTGGCCCCGTGAGATTATCGAGTGCTCCATGAATGAGGACACCC GGGCTACGGGAAAGAGGTGGACATGTGGAGCACTGGCGTCATCATGTACACGCTGCTG GCCGGCTCCCCGCCCTTCTGGCACCGGAAGCAGATGCTGATGCTGAGGATGATCATGA GCGGCAACTACAGTTTGGCTCGCCCGAGTGGGATGATTACTCGGACACCGTGAAGGA CCTGGTCTCCCGATTCTGGTGGTGCAACCCAGAACCGCTACACAGCGGAAGAGGCC TTGGCACACCCCTTCTCCAGCAGTACTTGGTGGAGGAAGTGCAGCACTTCAGGCCCC GGGGGAAGTTCAAGGTGATCGCTCTGACCGTCTGGCTTCACTGCGGATCTACTACCA GTACCGCCGGGTGAAGCCTGTGACCCGGGAGATCGTCATCCGAGACCCCTATGCCCTC CGGCCTCTGCGCCGGCTCATCGACGCTACGCTTTCGGAATCTATGGCCACTGGGTGA AGAAGGGGCAGCAGCAGAACCGGGCAGCCCTTTTCGAGAACACACCAAGGCCGTGCT CCTCTCCCTGGCCGAGGAGACTACTGAGGGGCT		
	ORF Start: ATG at 83		ORF Stop: TGA at 1244
	SEQ ID NO: 58	387 aa	MW at 45023.3kD
NOV2k, CG101996-03 Protein Sequence	MTRDEALPDSHSAQDFYENYEPKEILGRGVSSVVRRCIHKPTSQYAVKVIDVTGGGS FSPEEVRELREATLKEVDILRKVSGHPNIIQLKDTYETNTFFFLVFDLMKRGELFDYL TEKVTLSKETRKIMRALLEVICTLHKLNIHVHDLKPENILLDDNMNIKLTDGFGSCQ LEPGERLREVCCTPSYLAPEIIECMNEHDHPGYGKEVDMWSTGVIMYTLLAGSPPFWR RKQMLMLRMIMSGNYQFGSPWDDYSDTVKDLVSRFLVVQPQNRITYAEALAHPPFQQ YLVEEVVRFHSPRGKFKVIALTVLASVRIYYQYRRVKPVTREIVIRDPYALRPLRLRID AYAFRIYGHVVKKGQQONRAALFENTPKAVLLSLAEEDY		
	SEQ ID NO: 59	1194 bp	
NOV2l, CG101996-05	CGCGGATCCACCATGACCCGGGACGAGGCACTGCCGGACTCTCATTCTGCACAGGACT TCTATGAGAATTATGAGCCCAAAGAGATCTGGGCAGGGCGGTAGCAGTGTGGTCAG GCGATGCATCCACAAGCCCACGAGCCAGGAGTACGCCGTGAAGGTATCGACGTCACC		

DNA Sequence	GGTGGAGGCAGCTTCAGCCCGGAGGAGGTGCGGGAGCTGCGAGAAGCCACGCTGAAGGAGGTGGACATCCTGCGCAAGGTCTCAGGGCACCCCAACATCATACAGCTGAAGGACAC TTATGAGACCAACACTTTTCTTCTTCTTGGTGTTTGACCTGATGAAGAGAGGGGAGCTC TTTGACTACCTCACTGAGAAGGTACCTTGAGTGAGAAGGAAACCAGAAAGATCATGC GAGCTCTGCTGGAGGTGATCTGCACCTTGCACAACTCAACATCGTGCACCCGGGACCT GAAGCCCCGAGAACATTCTCTTGGATGACAACATGAACATCAAGCTCACAGACTTTGGC TTTTCTGCCAGCTGGAGCCGGGAGAGAGGCTGCGAGAGGTCTGCGGGACCCCCAGTT ACCTGGCCCCTGAGATTATCGAGTGCTCCATGAATGAGGACCACCCGGGCTACGGGAA AGAGGTGGACATGTGGAGCACTGGCGTCATCATGTACACGCTGCTGGCCGGCTCCCCG CCCTTCTGGCACCCGAAGCAGATGCTGATGCTGAGGATGATCATGAGCGGCAACTACC AGTTTGGCTCGCCGAGTGGGATGATTACTCGGACACCGTGAAGGACCTGGTCTCCCG ATTCTGGTGGTGCAACCCAGAACCGCTACACAGCGGAAGAGGCCTTGGCACACCC TTTCTCCAGCAGTACTTGGTGGAGGAAGTGCGGCACCTTCAGCCCCGGGGGAAGTTCA AGGTGATCGCTCTGACCGTGTCTGGCTTCAGTGCGGATCTACTACCAGTACCGCCGGGT GAAGCCTGTGACCCGGGAGATCGTCATCCGAGACCCCTATGCCCTCCGGCCTCTGCGC CGGCTCATCGACGCTACGCTTTCGAATCTATGGCCACTGGGTGAAGAAGGGGCAGC AGCAGAACCCGGGCAGCCCTTTTCGAGAACACACCCAAGGCCGTGCTCCTCTCCCTGGC CGAGGAGGACTACTGAGCGGCCGCTTTTTTCTCT		
	ORF Start: at 1		ORF Stop: TGA at 1174
	SEQ ID NO: 60	391 aa	MW at 45424.7kD
NOV2l, CG101996-05 Protein Sequence	RGSTMTRDEALPDSHAQDFYENYEPKEILGRGVSSVVRRCIHKPTSQEYAVKVIDVT GGGSFSPPEVRELREATLKEVDILRKVSGHPNIIQLKDTYETNTFFFLVFDLMKRGE FDYLTEKVTLSKETRKIMRALLEVICTLHKLNIVHRDLKPENILLDDNMNIKLTDFG FSCQLEPGERLREVCCTPSYLAPEIIECSMNEDHPGYGKEVDMWSTGVIMYTLLAGSP PFWHRKQMLMLRMIMSGNYQFGSPWDDYSDTVKDLVSRFLVVQPQNRITYAEALAHF FFQQYLVVEVRHFSRPGKFKVIALTVLASVRIYYQYRRVKPVTREIVIRDPYALRPLR RLIDAYAFRIYGHVWKKGQQQNRALFENTPKAVLLSLAEDDY		
	SEQ ID NO: 61	1165 bp	
NOV2m, CG101996-06 DNA Sequence	CATGACCCGGGACGAGGCACCTGCGGACTCTCATTCTGCACAGGACTTCTATGAGAAT TATGAGCCCAAGAGATCCTGGGCAGGGGCGTTAGCAGTGTGGTCAGGCGATGCATCC ACAAGCCACGAGCCAGGAGTACGCCGTGAAGGTCACTCGACGTACCCGGTGGAGGCAG CTTCAGCCCGGAGGAGGTGCGGGAGCTGCGAGAAGCCACGCTGAAGGAGGTGGACATC CTGCGCAAGGTCTCAGGGCACCCCAACATCATACAGCTGAAGGACACTTATGAGACCA ACACTTTCTTCTTCTTGGTGTGTGACCTGATGAAGAGAGGGGAGCTCTTTGACTACCT CACTGAGAAGGTACCTTGAGTGAGAAGGAACCAAGAAAGATCATGCGAGCTCTGCTG GAGGTGATCTGCACCTTGCACAACTCAACATCGTGCACCGGGAACCTGAAGCCCGAGA ACATTCTCTTGGATGACAACATGAACATCAAGCTCACAGACTTTGGCTTTTCTGCCA GCTGGAGCCGGGAGAGAGGCTGCGAGAGGTCTGCGGGACCCCAAGTTACCTGGCCCC GAGATTATCGAGTGCTCCATGAATGAGGACCACCCGGGCTACGGGAAAGAGGTGGACA TGTGGAGCACTGGCGTCATCATGTACACGCTGCTGGCCGGCTCCCCGCCCTTCTGGCA CCGGAAGCAGATGCTGATGCTGAGGATGATCATGAGCGGCAACTACCAGTTTGCTCG CCCGAGTGGGATGATTACTCGGACACCGTGAAGGACCTGGTCTCCCGATTCTGGTGG TGCAACCCCGAACCCTACACAGCGGAAGAGGCCTTGGCACACCCCTTCTTCCAGCA GTACTTGGTGGAGGAAGTGCAGCACTTCAGCCCCGGGGGAAGTTCAAGGTGATCGCT CTGACCGTGTGGCTTCAGTGGGATCTACTACCAGTACCGCCGGGTGAAGCCTGTGA CCCGGGAGATCGTCATCCGAGACCCCTATGCCCTCCGGCCTCTGCGCCGGCTCATCGA CGCCTACGCTTTCGAATCTATGGCCACTGGGTGAAGAAGGGGCAGCAGCAGAACCGG GCAGCCCTTTTCGAGAACACACCCAAGGCCGTGCTCCTCTCCCTGGCCGAGGAGACT ACTGA		
	ORF Start: ATG at 2		ORF Stop: TGA at 1163
	SEQ ID NO: 62	387 aa	MW at 45023.3kD
NOV2m, CG101996-06 Protein Sequence	MTRDEALPDSHAQDFYENYEPKEILGRGVSSVVRRCIHKPTSQEYAVKVIDVTGGGS FSPPEVRELREATLKEVDILRKVSGHPNIIQLKDTYETNTFFFLVFDLMKRGE FDYL TEKVTLSKETRKIMRALLEVICTLHKLNIVHRDLKPENILLDDNMNIKLTDFGFSQ LEPGERLREVCCTPSYLAPEIIECSMNEDHPGYGKEVDMWSTGVIMYTLLAGSPFWH RKQMLMLRMIMSGNYQFGSPWDDYSDTVKDLVSRFLVVQPQNRITYAEALAHFPPQQ YLVVEVRHFSRPGKFKVIALTVLASVRIYYQYRRVKPVTREIVIRDPYALRPLRLID		

	AYAFRIYGHVWKKGQQQNRAALFENTPKAVLLSLAEEDY		
	SEQ ID NO: 63	927 bp	
NOV2n, CG101996-07 DNA Sequence	ACCATGGGACATCATCACCACCATCACACCCGGGACGAGGCACTGCCGGACTCTCATT CTGCACAGGACTTCTATGAGAATTATGAGCCCAAAGAGATCCTGGGCAGGGGCGTTAG CAGTGTGGTCAGGCGATGCATCCACAAGCCACGAGCCAGGAGTACGCCGTGAAGGTC ATCGACGTCACCGGTGGAGGCAGCTTCAGCCCGGAGGAGGTGCGGGAGCTGCGAGAAG CCACGCTGAAGGAGGTGGACATCCTGCGCAAGGTCTCAGGGCACCCCAACATCATACA GCTGAAGGACACTTATGAGACCAACACTTTCTTCTTCTTGGTGTGTTGACCTGATGAAG AGAGGGGAGCTCTTTGACTACCTCACTGAGAAGGTCACCTTGAGTGAGAAGGAAACCA GAAAGATCATGCGAGCTCTGCTGGAGGTGATCTGCACCTTGACAAACTCAACATCGT GCACCCGGACCTGAAGCCCGAGAACATTCTCTTGGATGACAACATGAACATCAAGCTC ACAGACTTTGGCTTTTCTGCGAGCTGGAGCCGGGAGAGAGGCTGCGAGAGGTCTGCG GGACCCCCAGTTACCTGGCCCCCTGAGATTATCGAGTGCTCCATGAATGAGGACCAACC GGGCTACGGGAAAGAGGTGGACATGTGGAGCACTGGCGTCATCATGTACACGCTGCTG GCCGGCTCCCCGCCCTTCTGGCACCGGAAGCAGATGCTGATGCTGAGGATGATCATGA GCGGCAACTACCAGTTTGGCTCGCCCGAGTGGGATGATTACTCGGACACCGTGAAGGA CCTGGTCTCCCGATTCTTGGTGGTGAACCCAGAACCGCTACACAGCGGAAGAGGCC TTGGCACACCCCTTCTTCAGCAGTACTTGGTGAGGAAGTGGCGCACTTCAGCTGA		
	ORF Start: at 1		ORF Stop: TGA at 925
	SEQ ID NO: 64	308 aa	MW at 35743.4kD
NOV2n, CG101996-07 Protein Sequence	TMGHHHHHTRDEALPDSHAQDFYENYEPKEILGRGVSSVVRRCIHKPTSQEYAVKV IDVTGGGSFSPEEVRELREATLKEVDILRKVSGHPNIIQLKDTYETNTFFFLVFDLMK RGELFDYLTEKVTLSKETRKIMRALLEVICTLHKLNIVHRDLKPENILLDDNMNIKL TDFGFSCQLEPGERLREVCGPSYLAPEIIECSMNEDHPGYGKEVDMWSTGVIMYTL AGSPFFWHRKQMLMLRMIMSGNYQFGSPWDDYSDTVKDLVSRFLVVQPQNRYTAEBA LAHPFFQOYLVEEVVRFHS		
	SEQ ID NO: 65	924 bp	
NOV2o, CG101996-08 DNA Sequence	ACCATGACCCGGGACGAGGCACTGCCGGACTCTCATTCTGCACAGGACTTCTATGAGA ATTATGAGCCCAAAGAGATCCTGGGCAGGGGCGTTAGCAGTGTGGTCAGGCGATGCAT CCACAAGCCCAAGAGGAGGAGTACGCCGTGAAGGTATCGAGCTACCGGTGGAGGC AGCTTCAGCCCGGAGGAGGTGCGGGAGCTGCGAGAAGCCACGCTGAAGGAGGTGGACA TCCTGCGCAAGGTCTCAGGGCACCCCAACATCATACAGCTGAAGGACACTTATGAGAC CAACACTTTCTTCTTCTTGGTGTGTTGACCTGATGAAGAGAGGGGAGCTCTTTGACTAC CTCCTGAGAAGGTCACCTTGAGTGAGAAGGAAACCAGAAAGATCATGCGAGCTCTGC TGGAGGTGATCTGCACCTTGACAAACTCAACATCGTGCACCGGACCTGAAGCCCGA GAACATTCTCTTGGATGACAACATGAACATCAAGCTCACAGACTTTGGCTTTTCCTGC CAGCTGGAGCCGGGAGAGAGGCTGCGAGAGGTCTGCGGGACCCCACTTACCTGGCCC CTGAGATTATCGAGTGCTCCATGAATGAGGACCAACCGGGCTACGGGAAAGAGGTGGA CATGTGAGCACTGGCGTCATCATGTACACGCTGCTGGCCGGCTCCCCGCCCTTCTG CACCGAAGCAGATGCTGATGCTGAGGATGATCATGAGCGGCAACTACCAGTTTGGCT CGCCCGAGTGGGATGATTACTCGGACACCGTGAAGGACCTGGTCTCCCGATTCTTGGT GGTGCAACCCAGAACCGCTACACAGCGGAAGAGGCCTTGGCACACCCCTTCTTCCAG CAGTACTTGGTGGAGGAAGTGCAGCACTTCAGCCATCATCACCACCATCACTGA		
	ORF Start: at 1		ORF Stop: TGA at 922
	SEQ ID NO: 66	307 aa	MW at 35686.3kD
NOV2o, CG101996-08 Protein Sequence	TMTRDEALPDSHAQDFYENYEPKEILGRGVSSVVRRCIHKPTSQEYAVKVIDVTGGG SFSPEEVRELREATLKEVDILRKVSGHPNIIQLKDTYETNTFFFLVFDLMKRGELFDY LTEKVTLSKETRKIMRALLEVICTLHKLNIVHRDLKPENILLDDNMNIKLTDGFGSC QLEPGERLREVCGPSYLAPEIIECSMNEDHPGYGKEVDMWSTGVIMYTLLAGSPFFW HRKQMLMLRMIMSGNYQFGSPWDDYSDTVKDLVSRFLVVQPQNRYTAEBA LAHPFFQOYLVEEVVRFHSHHHHH		
	SEQ ID NO: 67	939 bp	
NOV2p, CG101996-09 DNA Sequence	CATATGACCCGGGACGAGGCACTGCCGGACTCTCATTCTGCACAGGACTTCTATGAGA ATTATGAGCCCAAAGAGATCCTGGGCAGGGGCGTTAGCAGTGTGGTCAGGCGATGCAT CCACAAGCCCAAGAGGAGGAGTACGCCGTGAAGGTATCGAGCTACCGGTGGAGGC ACCCTGAGCCCGGAGGAGGAGTACGCCGTGAAGGTATCGAGCTACCGGTGGAGGC		

	AGCTTCAGCCCGGAGGAGGTGCGGGAGCTGCGAGAAGCCACGCTGAAGGAGGTGGACA TCCTGCGCAAGGTCTCAGGGCACCCCAACATCATACAGCTGAAGGACACTTATGAGAC CAACACTTTCTTCTTCTTGGTGTTTGACCTGATGAAGAGAGGGGAGCTCTTTGACTAC CTCACTGAGAAGGTCACCTTGAGTGAGAAGGAAACCAGAAAGATCATGCGAGCTCTGC TGGAGGTGATCTGCACCTTGACAAACTCAACATCGTGACCCGGGACCTGAAGCCCGA GAACATTCTCTTGGATGACAACATGAACATCAAGCTCACAGACTTTGGCTTTTCTGTC CAGCTGGAGCCGGGAGAGAGGCTGCGAGAGGTCTGCGGGACCCCAAGTTACCTGGCCC CTGAGATTATCGAGTGCTCCATGAATGAGGACCACCCGGGCTACGGGAAAGAGGTGGA CATGTGGAGCACTGGCGTCATCATGTACGCTGCTGGCCGGCTCCCGCCCTTCTGG CACC CGAAGCAGATGCTGATGCTGAGGATGATCATGAGCGGCACTACCAGTTTGGCT CGCCCGAGTGGGATGATTACTCGGACACCGTGAAGGACCTGGTCTCCCGATTCTCTGGT GGTGCAACCCAGAACCGCTACACAGCGGAAGAGGCTTGGCACACCCCTTCTCCAG CAGTACTTGGTGGAGGAAGTGCGGCACCTTCAGCTGAGCGGCCGCACTCGAGCACCACC ACCACCACCAC		
	ORF Start: at 1		ORF Stop: TGA at 904
	SEQ ID NO: 68	301 aa	MW at 34899.5kD
NOV2p, CG101996-09 Protein Sequence	HMTREALPDHSAQDFYENYEPKEILGRGVSSVVRRCIHKPTSQEYAVKVIDVTGGG SFSPEEVRELREATLKEVDILRKVSGHPNIIQLKDTYETNTFFFLVFDLMKRGELFDY LTEKVTLSEKETRKIMRALLEVICTLHKLNIHVHRLKPENILLDDNMNIIKLTDGFGSC QLEPGERLREVCCTPSYLAPEIIIECSMNEDHPGYGKEVDMWSTGVIMYTLLAGSPFFW HRKQMLMLRMIMSGNYQFGSPEWDDYSDTVKDLVSRFLVVQPQNRYTAEELAHPPFFQ QYLVEEVVRHFS		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 2B.

Table 2B. Comparison of NOV2a against NOV2b through NOV2p.		
Protein Sequence	NOV2a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV2b	1..152 281..432	152/152 (100%) 152/152 (100%)
NOV2c	1..152 236..387	152/152 (100%) 152/152 (100%)
NOV2d	1..152 236..387	152/152 (100%) 152/152 (100%)
NOV2e	1..152 281..432	152/152 (100%) 152/152 (100%)
NOV2f	1..65 244..308	65/65 (100%) 65/65 (100%)
NOV2g	1..152 240..391	152/152 (100%) 152/152 (100%)
NOV2h	1..65 241..305	65/65 (100%) 65/65 (100%)
NOV2i	1..65 237..301	65/65 (100%) 65/65 (100%)
NOV2j	1..65	65/65 (100%)

	235..299	65/65 (100%)
NOV2k	1..152 236..387	152/152 (100%) 152/152 (100%)
NOV2l	1..152 240..391	152/152 (100%) 152/152 (100%)
NOV2m	1..152 236..387	152/152 (100%) 152/152 (100%)
NOV2n	1..65 244..308	65/65 (100%) 65/65 (100%)
NOV2o	1..65 237..301	65/65 (100%) 65/65 (100%)
NOV2p	1..65 237..301	65/65 (100%) 65/65 (100%)

Further analysis of the NOV2a protein yielded the following properties shown in Table 2C.

Table 2C. Protein Sequence Properties NOV2a	
PSort analysis:	0.5098 probability located in microbody (peroxisome); 0.4500 probability located in cytoplasm; 0.3051 probability located in lysosome (lumen); 0.1000 probability located in mitochondrial matrix space
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV2a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 2D.

Table 2D. Geneseq Results for NOV2a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV2a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
ABB09290	Human phosphorylase kinase gamma 2 (PHKG2) protein SEQ ID NO:4 - Homo sapiens, 406 aa. [WO200194365-A2, 13-DEC-2001]	1..140 239..378	82/140 (58%) 105/140 (74%)	5e-43
AAY43921	Rabbit protein kinase #3 - Oryctolagus cuniculus, 268 aa. [US5958784-A, 28-SEP-1999]	1..56 213..268	55/56 (98%) 55/56 (98%)	2e-26
AAY43922	Mouse protein kinase #3 - Mus sp, 268 aa. [US5958784-A, 28-SEP-1999]	1..56 213..268	50/56 (89%) 53/56 (94%)	2e-23

	1999]			
ABG10311	Novel human diagnostic protein #10302 - Homo sapiens, 886 aa. [WO200175067-A2, 11-OCT-2001]	44..140 615..718	49/104 (47%) 69/104 (66%)	1e-19
ABB58577	Drosophila melanogaster polypeptide SEQ ID NO 2523 - Drosophila melanogaster, 560 aa. [WO200171042-A2, 27-SEP-2001]	64..147 470..553	43/84 (51%) 57/84 (67%)	4e-17

In a BLAST search of public sequence databases, the NOV2a protein was found to have homology to the proteins shown in the BLASTP data in Table 2E.

Table 2E. Public BLASTP Results for NOV2a				
Protein Accession Number	Protein/Organism/Length	NOV2a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q16816	Phosphorylase B kinase gamma catalytic chain, skeletal muscle isoform (EC 2.7.1.38) (Phosphorylase kinase gamma subunit 1) - Homo sapiens (Human), 386 aa.	1..152 235..386	152/152 (100%) 152/152 (100%)	5e-84
KIRBFG	phosphorylase kinase (EC 2.7.1.38) catalytic chain, skeletal muscle - rabbit, 387 aa.	1..152 236..387	147/152 (96%) 149/152 (97%)	1e-81
P00518	Phosphorylase B kinase gamma catalytic chain, skeletal muscle isoform (EC 2.7.1.38) (Phosphorylase kinase gamma subunit 1) - Oryctolagus cuniculus (Rabbit), 386 aa.	1..152 235..386	147/152 (96%) 149/152 (97%)	1e-81
S00731	phosphorylase kinase (EC 2.7.1.38) catalytic chain [similarity] - rat, 388 aa.	1..151 236..386	142/151 (94%) 147/151 (97%)	3e-78
P13286	Phosphorylase B kinase gamma catalytic chain, skeletal muscle isoform (EC 2.7.1.38) (Phosphorylase kinase gamma subunit 1) - Rattus norvegicus (Rat), 387 aa.	1..151 235..385	142/151 (94%) 147/151 (97%)	3e-78

PFam analysis predicts that the NOV2a protein contains the domains shown in the Table 2F.

Table 2F. Domain Analysis of NOV2a			
Pfam Domain	NOV2a Match Region	Identities/ Similarities for the Matched Region	Expect Value
pkinese	3..53	16/54 (30%) 43/54 (80%)	4.4e-09

### Example 3.

The NOV3 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 3A.

Table 3A. NOV3 Sequence Analysis			
	SEQ ID NO: 69	2727 bp	
NOV3a, CG102822-01 DNA Sequence	AGAAGAGCGGAGCTGTGAGCAGTACTGCGGCCTCCTCTCCTCTCCTAACCTCGCTCTC GCGGCCTAGCTTTACCCGCCCGCTGCTCGGCGACCAGAACACCTTCCACCATGACCA CCTCAGCAAGTTCCCACTTAAATAAAGGCATCAAGCAGGTGTACATGTCCCTGCCTCA GGGTGAGAAAGTCCAGGCCATGTATATCTGGATCGATGGTACTGGAGAAGGACTGCGC TGCAAGACCCGACCTGGACAGTGTAGCCCAAGTGTGTGGAAGAGTTGCCCTGAGTGGG ATTTTCGATGGCTCTAGTACTTTACAGTCTGAGGGTTCCAACAGTGACATGTATCTCGT GCCTGCTGCCATGTTTCGGGACCCCTTCCGTAAGGACCCCTAACAAAGCTGGTGTATGT GAAGTTTCAAGTACAATCGAAGGCCCTGCAGAGACCAATTGAGGCACACCTGTAAAC GGATAATGGACATGGTGAGCAACCAGCACCCCTGGTTTGGCATGGAGCAGGAGTATAC CCTCATGGGGACAGATGGGCACCCCTTGGTTGGCCTTCCAACGGCTTCCCAGGGCCC CAGGGTCCATATTACTGTGGTGTGGGAGCAGACAGAGCCTATGGCAGGGACATCGTGG AGGCCCATTACCGGCCCTGCTTGTATGCTGGAGTCAAGATTGCGGGGACTAATGCCGA GGTCAATGCTGCCAGTGGGAATTCAGATTGGACCTTGTGAAGGAATCAGCATGGGA GATCATCTCTGGGTGGCCCGTTTCATCTTGCATCGTGTGTGTGAAGACTTTGGAGTGA TAGCAACCTTTGATCCTAAGCCATTCTTGGGAATGGAATGGTGCAGGCTGCCATAC CAACCTCAGCACCAAGGCCATCGGGGAGGAGAATGGTCTGAAGTACATCGAGGAGGCC ATTGAGAACTAAGCAAGCGGCACCAAGTACCACATCCGTGCCTATGATCCCAAGGGAG GCCTGGACAATGCCGACGCTTAAGTGGATTCCATGAAACCTCCAACATCAACGACTT TTCTGCTGGTGTAGCCAATCGTAGCGCCAGACTACGCATTTCCCGGACTGTTGGCCAG GAGAAGAAGGGTTACTTTGAAGATCGTCGCCCTCTGCCAACTGCGAGCCCTTTTCGG TGACAGAAGCCCTCATCCGACGCTGTCTTCAATGAAACCGGCATGAGCCCTTCCA GTACAAAAATTAAGTGGACTAGACCTCCAGCTGTTGAGCCCTCCTAGTTCTTTCATCC CTGACTCCAACCTTCCCCCTCTCCAGTTGTCCCGATTGTAAGTCAAGGGGTGGAAT ATCAAGGTCGTTTTTTTTCATTCCATGTGCCAGTTAATCTTGCTTTCTTTTGGTTGGC TGGGATAGAGGGGTCAAGTTATTAATTTCTTACACCTACCCTCCTTTTTCCTAT CACTGAAGCTTTTGTAGTCATTAGTGGGGAGGAGGGTGGGGAGACATAACCACTGCTT CCATTTAATGGGGTGACCTGTCCAATAGGCGTACGTATCCGGACAGAGCACGTTTGC AGAGGGGTCTCTCTCCAGGTAGCTGAAAGGGAAGACCTGACGTACTCTGGTTAGGTTA GGACTTGCCCTCGTGGTGGAACTTTTCTTAAAAAGTTATAACCAACTTTTCTATTAA AAGTGGGAATTAGGAGAGAAGGTAGGGGTTGGGAATCAGAGAGAATGGCTTTGGTCTC TTGCTTGTGGGACTAGCCTGGCTTGGGACTAAATGCCCTGCTCTGAACACAAGCTTAG TATAAACTGATGGATATCCCTACCTTGAAAGAAGAAAGGTTCTTACTGCTTGGTCTT TGATTTATCACACAAAGCAGAATAGTATTTTATATTTAAATGTAAGACAAAAAAT ATATGTATGGTTTGTGGATTATGTGTGTTTGGCTAAAGGAAAAAACCATCCAGGTC ACGGGGCACCAAAATTTGAGACAAATAGTCGGATTAGAAATAAAGCATCTCATTTGAG TAGAGAGCAAGGAAGTGGTTCTTAGATGGTGATCTGGGATTAGGCCCTCAAGACCCCT TTTGGGTTTCTGCCCTGCCACCCCTCTGGAGAAGGTGGCACTGATTAGTTAACAGACC AACACCGTTACTAGCAGTCACTGATCTCCGTGGCTTTGGTTTAAAGACACACTTGTC CACATAGGTTTAGAGATAAGAGTTGGCTGGTCAACTTGAGCATGTTACTGACAGAGGG GGTATTGGGGTTATTTCTGGTAGGAATAGCATGTCACTAAAGCAGGCCCTTGATATT AAATTTTTTAAAAAGCAAAATATAGAAGTTTAGATTTTAATCAAATTTGTAGGGTTT		

	CTAGGTATTACAGATGCTGTTGCTCAACGTCTCTACCTCTGCTCTGAGAGATGGGA CAGGCTGAGTCAAACACTGTAATTTTGTATCTTGATGTCTTTGTTAAGACTGCTGAAG AATTATTTTCTTTTATAATAAGGAATAAACCCACCTTTATTCCTTCATTTCATCT ACCATTTTCTGGTTCTGTGTTGGCTGTGGCAGGCCAGCTGTGGTTTTCTTTTGCCAT GACAACTTCTAATTGCCATGTACAGTATGTTCAAAGTCAAATAACTCCTCATTGTAAA CAAACGTGTAACGCCCAAAGCAGCACTTATAAATCAGCCTAACATAAAAAAAAAA A		
	ORF Start: at 68		ORF Stop: TAA at 1229
	SEQ ID NO: 70	387 aa	MW at 43593.8kD
NOV3a, CG102822-01 Protein Sequence	LYPPACSA TRTPSTMTSASSHLNKGIKQVYMSLPQGEKVQAMYIWIIDGTGEGLRCKT RTL DSEPKVEELPEWNFDGSSTLQSEGSNSDMYLVPAAMFRDPFRKDPNKLVLCEVF KYNRRPAETNLRHCTKRIMDMVSNQHPWFGMEQEYTLMGTDGHPFGWPSNGFPGPQGP YYCGVGADRAYGRDIVEAHYRACLYAGVKIAGTNAEVMPAQWFEQIGPCEGISMGDHL WVARFILHRVCEDFGVIATFDPKPIGNWNGAGCHTNFSTKAMREENGLKYIEEAIEK LSKRHQYHIRAYDPKGLDNARRLTGFHETSNINDFSAGVANRSARLRIPTVQGEKK GYFEDRRPSANCEPFSVTEALIRTCLLNETGDEPFQYKN		
	SEQ ID NO: 71	1366 bp	
NOV3b, CG102822-03 DNA Sequence	CGCGAGAGCAGGTTAGGAGAGGAGAGGAGGCCGAGTACTGCTCACACGCTCCGCTCT TCTCCCACTCTCGGCCTACCTTTACCCGCCCGCCTGCTCGGCAGCAGAACACCTTCC ACCATGACCACCTCAGCAAGTTCCCACTTAAATAAAGGCATCAAGCAGGTGTACATGT CCCTGCCTCAGGGTGAGAAAGTCCAGGCCATGTATATCTGGATCGATGGTACTGGAGA AGGACTGCGCTGCAAGACCCGGACCTGGACAGTGAGCCAAAGTGTGTGAAGAGTTG CCTGAGTGGAATTTGATGGCTCCAGTACTTTACAGTCTGAGGGTTCCAACAGTGACA TGTATCTCGTGCCTGCTGCCATGTTTCGGGACCCCTTCCGTAAGGACCCTAACAAGCT GGTGTATGTGAAGTTTTCAGTACAATCGAAGGCCTGCAGAGACCAATTTGAGGCAC ACCTGTAAACGGATAATGGACATGGTGAGCAACCAGCACCCCTGGTTTGGCATGGAGC AGGAGTATACCTCATGGGGACAGATGGGCACCCCTTTGGTTGGCCTTCCAACGGCTT CCCAGGGCCCCAGGTTCCATATTACTGTGGTGTGGGAGCAGACAGACCTATGGCAGG GACATCGTGGAGGCCATTACCGGGCCTGCTTGTATGCTGGAGTCAAGATTGCGGGGA CTAATGCCGAGGTCATGCCTGCCAGTGGGAATTTAGATTGGACCTTGTGAAGGAAT CAGCATGGGAGATCATCTCTGGGTGGCCCCGTTTCATCTGCATCGTGTGTGAAGAC TTTGGAGTGATAGCAACCTTTGATCCTAAGCCCATTCCTGGGAAC TGAATGGTGCAG GCTGCCATACCAACTTCAGCACCAAGGCCATGCGGGAGGAGAATGGTCTGAAGTACAT CGAGGAGGCCATTGAGAACTAAGCAAGCGGCACCAAGTACCACATCCGTGCCTATGAT CCCAAGGGAGGCCTGGACAATGCCCGACGTCTAACTGGATTCCATGAAACCTCCAACA TCAACGACTTTTCTGGTGGTGTAGCCAATCGTAGCGCCAGCATACGATTCCCCGGAC TGTTGGCCAGGAGAAGAAGGGTTACTTTGAAGATCGTCGCCCCCTCTGCCAATGCGAC CCCTTTTCGGTGACAGAAGCCCTCATCCGACGTGTCTTCTCAATGAACCGCGGATG AGCCCTTCCAGTACAAAATTAAGTGGACTAGACCTCCAGCTGTTGAGCCCCCTCTAG TTCTTCATCCCACTCCAACCTTCCCCCTCTCCAGTTGTCCCGATTGTAAC TCAAG GGTGAATATCAAGGTCGTTTTTTTCATTC		
	ORF Start: ATG at 120		ORF Stop: TAA at 1239
	SEQ ID NO: 72	373 aa	MW at 42050.0kD
NOV3b, CG102822-03 Protein Sequence	MTSASSHLNKGIKQVYMSLPQGEKVQAMYIWIIDGTGEGLRCKRTL DSEPKVEELP EWNFDGSSTLQSEGSNSDMYLVPAAMFRDPFRKDPNKLVLCEVFKYNRRPAETNLRH CKRIMDMVSNQHPWFGMEQEYTLMGTDGHPFGWPSNGFPGPQGPYYCGVGADRAYGR DIVEAHYRACLYAGVKIAGTNAEVMPAQWFEQIGPCEGISMGDHLWVARFILHRVCEDF GVIATFDPKPIGNWNGAGCHTNFSTKAMREENGLKYIEEAIEKLSKRHQYHIRAYDP KGLDNARRLTGFHETSNINDFSGGVANRSASIRIPTVQGEKKGYFEDRRPSANCDP FSVTEALIRTCLLNETGDEPFQYKN		
	SEQ ID NO: 73	2631 bp	
NOV3c, CG102822-02 DNA Sequence	ATGACCACCTCAGCAAGTTCCTCACTTAAATAAAGGCATCAAGCAGGTGTACATGTCCC TGCTCAGGGTGAGAAAGTCCAGGCCATGTATATCTGGATCGATGGTACTGGAGAAGG ACTGCGCTGCAAGACCCGACCTTGGACAGTGAGCCAAAGTGTGTGGAAAGATTGCTC GAGTGGAAATTCGATGGCTCTAGTACTTTACAGTCTGAGGGTTCCAACAGTGACATGT ATCTCGTGCCTGCTGCCATGTTTCGGGACCCCTTCCGTAAGGACCCTAACAAGCTGGT		

	GTTATGTGAAGTTTTCAGTACAATCGAAGGCCTGCAGAGACCAATTTGAGGCACACC TGTAAACGGATAATGGACATGGTGAGCAACCAGCACCCCTGGTTTGGCATGGAGCAGG AGTATACCCTCATGGGGACAGATGGGCACCCCTTTGGTTGGCCCTTCCAACGGCTTCCC AGGGCCCCAGGGTCCATATTAAGTGTGGTGGGAGCAGACAGAGCCTATGGCAGGGAC ATCGTGGAGGCCCATTAACGGGCCCTGCTTGTATGTGGAGTCAAGATTGCGGGGACTA ATGCCGAGGTCATGCCCTGCCAGTGGGAATTTTCAGATTGGACCTTGTGAAGGAATCAG CATGGGAGATCATCTCTGGGTGGCCCGTTTCATCTTGCATCGTGTGTGTGAAGACTTT GGAGTGATAGCAACCTTTGATCCTAAGCCCATTCCTGGGAACGGAATGGTGCAGGCT GCCATACCAACTTCAGCACCAAGGCCATGCGGGAGGAGAATGGTCTGAAGTACATCGA GGAGGCCATTGAGAACTAAGCAAGCGGCACCAAGTACCACATCCGTGCCCTATGATCCC AAGGGAGGCCCTGGACAAATGCCCGACGTCTAAGTGGATTCCATGAAACCTCCAACATCA ACGACTTTTCTGCTGGTGTAGCCAATCGTAGCGCCAGCATAACGATATCCCCGACTGT TGGCCAGGAGAAGAAGGGTTACTTTGAAGATCGTCGCCCTCTGCCAATCGGACACCCC TTTTCGGTGACAGAAGCCCTCATCCGCACGTGTCTTCTCAATGAAACCGGCGATGAGC CCTTCCAGTACAAAAATTAAGTGGACTAGACCTCCAGCTGTTGAGCCCCCTCTAGTTC TTCATCCCCTCCAACCTCTTCCCCCTCTCCAGTTGTCCCGATTGTAACCTCAAAGGGT GGAATATCAAGGTCGTTTTCATTCCATGTGCCAGTTAATCTTGCTTTCTTTGTT TGGCTGGGATAGAGGGGTCAAGTTATTAATTTCTTACACCTACCCCTCTTTTTCCTC CTATCACTGAAGCTTTTGTAGTGCATTAGTGGGGAGGAGGGTGGGGAGACATAACCACT GCTTCCATTAAATGGGGTGCACCTGTCCAATAGGCGTAGCTATCCGGACAGAGCACGT TTGCAGAAGGGGTCTCTTCTTCCAGGTAGCTGAAAGGGGAAGACCTGACGTACTCTG GTTAGGTTAGGACTTGCCCTCGTGGTGGAACTTTTCTTAAAGGTTATAACCAACTT TTCTATTAAAGTGGGAATTAGGAGAGAAGGTAGGGGTTGGGAATCAGAGAGAATGGC TTTGGTCTCTTGCTTGTGGGACTAGCCTGGCTTGGGACTAAATGCCCTGCTCTGAACA CGAAGCTTAGTATAAACTGATGGATATCCCTACCTGAAAGAAGAAAAGGTTCTTACT GCTTGGTCTTGATTATCACACAAAGCAGAATAGTATTTTTATATTTAAATGTAAAG ACAAAAACTATATGTATGGTTTTGTGGATTATGTGTGTTTTGCTAAAGGAAAAAACC ATCCAGGTCACGGGGCACCAAAATTTGAGACAAATAGTCGGATTAGAAATAAGCATCT CATTTTGAGTAGAGAGCAAGGGAAGTGGTTCTTAGATGGTGTCTGGGATTAGGCCCT CAAGACCTTTTGGGTTTCTGCCCTGCCACCCCTCTGGAGAAGGTGGGCACTGGATTAG TTAACAGACAACACGTTACTAGCAGTCACTTGATCTCCGTGGCTTTGGTTTAAAGAC ACACCTTGTCACATAGGTTTAGAGATAAGAGTTGGCTGGTCACTTGAGCATGTTACT GACAGAGGGGTATTGGGGTTATTTCTGGTAGGAATAGCATGCTACTAAAGCAGGCC TTTTGATATTAAATTTTAAAGCAAAATTTATAGAAGTTTAGATTTTAATCAAATT TGATGGGTTTCTAGGTAATTTTACAGAATTGCTTGTGTTGCTTCAACTGTCTCCTACC TCTGCCTCTTGGAGGAGATGGGACAGGGCTGGAGTCAAAACACTTGTAAATTTGTATC TTGATGCTTTGTAAAGACTGCTGAAGAATTATTTTTTTCTTTTATAATAAGGAATA AACCCACCTTTATTCCTTCATTTTCATCTACCATTTCTGGTTCTGTGTGTTGGCTGTG GCAGGCCAGCTGTGGTTTCTTTTGGCATGACAACCTCTAATTGCCATGTACAGTATG TTCAAAGTCAAATAACTCCTCATTGTAAACAACTGTGTAACGCCCAAAGCAGCACT TATAAATCAGCCTAACATAAG		
	ORF Start: ATG at 1		ORF Stop: TAA at 1120
	SEQ ID NO: 74	373 aa	MW at 42064.0kD
NOV3c, CG102822-02 Protein Sequence	MTTSSASHLNKGIKQVYMSLPQGEKVQAMYIWDGTGEGLRCKTRTLTLDSEPKVEELP EWNFDGSSTLQSEGSNSDMLVPAAMFRDPFRKDPNKLVLCEVFKNRRPAETNLRHT CKRIMDMVSNQHPWFQMEQEYTLTGTDGHPFGWPSNGFPGPQGPYYCGVGADRAYGRD IVEAHYRACLYAGVKIAGTNAEVMQAQWFEQIGPCEGISMGDHLWVARFILHRVCEDF GVIATFDPKPIPGNWNAGCHTNFSTKAMREENGLKYIEEAEKLSKRHQYHIRAYDP KGGLDNARRLTGPHETSNINDFSAGVANRSASIRIPRTVQGEKKGYFEDRRPSANCDP FSVTEALIRTCLLNETGDEPFQYKN		
	SEQ ID NO: 75	2775 bp	
NOV3d, CG102822-04 DNA Sequence	GGCACGAGGGAAGAGCGGAGCGTGTGAGCAGTACTGCGGCCCTCCTCTCCTCTCCTAAC CTCGCTCTCGCGGCCTACCTTTACCCGCCCGCTGCTCGGCGACAGAACCTTCCA CCATGACCACCTCAGCAAGTTCCTCACTTAAATAAAGGCATCAAGCAGGTGTACATGTC CCTGCCTCAGGGTGAGAAAGTCCAGGCCATGTATATCTGGATCGATGGTACTGGAGAA GGACTGCGCTGCAAGACCCGACCTGGACAGTGAGCCCAAGTGTGTGGAAGAGTTGC CTGAGTGAATTTGATGGCTCCAGTACTTTACAGTCTGAGGGTTCCAACAGTGACAT GTATCTCGTGCCTGTGCCATGTTTCGGGACCCCTTCCGTAAGGACCTTAAACAAGCTG		

	GTGTTATGTGAAGTTTTCAGTACAATCGAAGGCCTGCAGAGACCAATTTGAGGCACA CCTGTAAACGGATAATGGACATGGTGAGCAACCAGCACCCCTGGTTTGGCATGGAGCA GGAGTATACCTCATGGGGACAGATGGGCACCCCTTTGGTTGGCCCTCCAACGGCTTC CCAGGGCCCCAGGGTCCATATTACTGTGGTGTGGGAGCAGACAGAGCCTATGGCAGGG ACATCGTGGAGGCCCATTTACCGGGCCCTGCTTGTATGCTGGAGTCAAGATTGCGGGGAC TAATGCCGAGGTGATGCCTGCCAGTGGGAATTTAGATTGGACCTTGTGAAGGAATC AGCATGGGAGATCATCTCTGGGTGGCCCGTTTCATCTTGCATCGTGTGTGTGAAGACT TTGGAGTGATAGCAACCTTTGATCCTAAGCCCATTCTCTGGGAACTGGAATGGTGCAGG CTGCCATACCAACTTCAGCACCAAGGCCATGCGGGAGGAGAATGGTCTGAAGTACATC GAGGAGGCCATTGAGAACTAAGCAAGCGGCACCAGTACCACATCCGTGCCTATGATC CCAAGGGAGGCCTGGACAATGCCGACGTCTAACTGGATTCCATGAAACCTCCAACAT CAACGACTTTTCTGCTGGTGTAGCCAATCGTAGCGCCAGCATAACGATCCCCGGACT GTTGGCCAGGAGAAGAAGGGTTACTTTGAAGATCGTCGCCCCCTGCGCAACTGCGACC CCTTTTCGGTGACAGAAGCCCTCATCCGCACGTGCTTCTCAATGAAACCGGCGATGA GCCCTTCCAGTACAAAAATTAAGTGGACTAGACCTCCAGCTGTTGAGCCCCCTCTAGT TCTTCTACCTCCACTCCAACCTTCTCCCCCTCTCCAGTTGTCCCGATTGTAACCTCAAGG GTGGAATATCAAGGTCGTTTTTTTTCATTCATGTGCCAGTTAATCTTGCTTTCTTTG TTTGGCTGGGATAGAGGGGTCAAGTTATTAATTTCTTCACACCTACCTCCTTTTTTTT CCCTATCACTGAAGCTTTTTAGTGCATTAGTGGGGAGGAGGGTGGGGAGACATAACCA CTGCTTCCATTTAATGGGGTGCACCTGTCCAATAGGCGTAGCTATCCGGACAGAGCAC GTTTGCAGAAGGGGGACTCTTCTTCCAGGTAGCTGAAAGGGGAAGACCTGACGTACTC TGGTTAGGTTAGGACTTGCCCTCGTGGTGGAAACTTTTCTTAAAAAGTTATAACCAAC TTTTCTATTAAAAAGTGGGAATTAGGAGAGAAGGTAGGGGTTGGGAATCAGAGAGAATC GCTTTGGTCTCTTGTCTTGTGGGACTAGCCTGGCTTGGGACTAAATGCCCTGCTCTGAA CACGAAGCTTAGTATAAACTGATGGATATCCCTACCTTGAAAGAAGAAAAGGTTCTTA CTGCTTGGTCCCTTGATTTATCACACAAAGCAGAATAGTATTTTATATTAAATGTAA AGACAAAAACTATATGATGGTTTTTGTGGATTATGTGTGTTTTGCTAAAGGAAAAAA CCATCCAGGTCACGGGGCACCAAATTTGAGACAAATAGTCGGATTAGAAATAAAGCAT CTCATTTTGTAGTAGAGCAAGGGAAGTGGTTCTTAGATGGTGATCTGGGATTAGGCC CTCAAGACCCTTTTGGGTTTCTGCCCTGCCACCCCTCTGGAGAAGGTGGGCACTGGAT TAGTTAACAGACGACACGTTACTAGCAGTCACTTGATCTCCGTGGCTTTGGTTTAAAA GACACACTTGTCCACATAGGTTTAGAGATAAGAGTTGGCTGGTCAACTTGAGCATGTT ACTGACAGAGGGGGTATTGGGGTTATTTTCTGGTAGGAATAGCATGTCACTAAAGCAG GCCTTTTGATATTAAATTTTTTAAAAAGCAAAATTATAGAAGTTAGATTTTAATCAA ATTTGTAGGGTTTCTAGGTAATTTTTACAGAATTGCTTGTTTGTCTCAACTGTCTCCT ACCTCTGCTCTTGGAGGAGATGGGGACAGGGCTGGAGTCAAAACACTTGTAATTTTGT ATCTTGATGCTTTGTTAAGACTGCTGAAGAATTATTTTTCTTTTATAATAAGGAA TAAACCCACCTTTATTCCTTCATTTTCATCTACCATTTTCTGGTTCTTGTGTTGGCTG TGGCAGGCCAGCTGTGGTTTTCTTTTGGCATGACAACTTCTAATTGCCATGTACAGTA TGTTCAAAGTCAAATAACTCCTCATTGTAACAACTGTGTAAGTGCCTCAAGCAGCA CTTATAAATCAGCTAACATAAGAAAAAAAAAAAAAAAAAAAAAAAAAAAAA		
	ORF Start: ATG at 119		ORF Stop: TAA at 1238
	SEQ ID NO: 76	373 aa	MW at 42064.0kD
NOV3d, CG102822-04 Protein Sequence	MTTSASHLNKGIKQVYMSLPQGEKVQAMYIWIWIDGTGEGLRCKRTLDSPEPKVEELP EWNFDGSSLTQSEGSNSDMLVPAAMFRDPFRKDPNKLVLCEVFKYNRRPAETNLRHT CKRIMDMVSNQHPWFGMEQEYTLMGTDGHPFGWPSNGFPGPQGPYYCGVGADRAYGRD IVEAHYRACLYAGVKIAGTNAEVMQAQWEPQIGPCEGISMGDHLWVARFILHRVCEDF GVIATFDPKPIPGNWNAGCHTNFSTKAMREENGLKYIEEAIEKLSKRHOYHIRAYDP KGGLDNARRLTGFHETSNINDFSAGVANRSASIRIPRTVGQEKKGYFEDRRPSANCDP FSVTEALIRTCLLNETGDEFPQYKN		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 3B.

Table 3B. Comparison of NOV3a against NOV3b through NOV3d.		
Protein Sequence	NOV3a Residues/	Identities/

	Match Residues	Similarities for the Matched Region
NOV3b	15..387 1..373	369/373 (98%) 371/373 (98%)
NOV3c	15..387 1..373	370/373 (99%) 372/373 (99%)
NOV3d	15..387 1..373	370/373 (99%) 372/373 (99%)

Further analysis of the NOV3a protein yielded the following properties shown in Table 3C.

Table 3C. Protein Sequence Properties NOV3a	
PSort analysis:	0.5025 probability located in mitochondrial matrix space; 0.4633 probability located in microbody (peroxisome); 0.2227 probability located in mitochondrial inner membrane; 0.2227 probability located in mitochondrial intermembrane space
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV3a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 3D.

Table 3D. Geneseq Results for NOV3a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV3a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAP70501	Chinese hamster glutamine synthetase gene product - Cricetulus griseus, 373 aa. [WO8704462-A, 30-JUL-1987]	15..387 1..373	347/373 (93%) 361/373 (96%)	0.0
ABG08130	Novel human diagnostic protein #8121 - Homo sapiens, 338 aa. [WO200175067-A2, 11-OCT-2001]	15..333 1..320	304/327 (92%) 305/327 (92%)	0.0
ABB58458	Drosophila melanogaster polypeptide SEQ ID NO. 2166 - Drosophila melanogaster, 369 aa. [WO200171042-A2, 27-SEP-2001]	18..377 9..369	235/361 (65%) 292/361 (80%)	e-150
ABB65740	Drosophila melanogaster polypeptide SEQ ID NO. 24012 - Drosophila melanogaster, 399 aa. [WO200171042-A2, 27-SEP-2001]	15..377 36..399	219/365 (60%) 271/365 (74%)	e-132

ABB59358	Drosophila melanogaster polypeptide SEQ ID NO 4866 - Drosophila melanogaster, 399 aa. [WO200171042-A2, 27-SEP-2001]	15..377 36..399	219/365 (60%) 271/365 (74%)	e-132
----------	---	--------------------	--------------------------------	-------

In a BLAST search of public sequence databases, the NOV3a protein was found to have homology to the proteins shown in the BLASTP data in Table 3E.

Table 3E. Public BLASTP Results for NOV3a				
Protein Accession Number	Protein/Organism/Length	NOV3a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
AJHUQ	glutamate--ammonia ligase (EC 6.3.1.2) - human, 373 aa.	15..387 1..373	372/373 (99%) 373/373 (99%)	0.0
P15104	Glutamine synthetase (EC 6.3.1.2) (Glutamate--ammonia ligase) - Homo sapiens (Human), 373 aa.	15..387 1..373	370/373 (99%) 372/373 (99%)	0.0
AAH31964	Similar to glutamine synthetase - Homo sapiens (Human), 373 aa.	15..387 1..373	368/373 (98%) 370/373 (98%)	0.0
P46410	Glutamine synthetase (EC 6.3.1.2) (Glutamate--ammonia ligase) - Sus scrofa (Pig), 373 aa.	15..387 1..373	357/373 (95%) 364/373 (96%)	0.0
Q91VC6	Glutamine synthetase (EC 6.3.1.2) (Hypothetical 42.1 kDa protein) - Mus musculus (Mouse), 373 aa.	15..387 1..373	350/373 (93%) 362/373 (96%)	0.0

PFam analysis predicts that the NOV3a protein contains the domains shown in the Table 3F.

Table 3F. Domain Analysis of NOV3a			
Pfam Domain	NOV3a Match Region	Identities/ Similarities for the Matched Region	Expect Value
gln-synt	38..366	133/375 (35%) 298/375 (79%)	3e-198

#### 5 Example 4.

The NOV4 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 4A.

Table 4A. NOV4 Sequence Analysis
----------------------------------

	SEQ ID NO: 77	1888 bp	
NOV4a, CG103241-01 DNA Sequence	AGCAGCCGGATGCCCGGGCCCACTGGGCGGGCCAGTGGCCGCTTGCGGGATGAGCAGA CTGCTGGGGGGGACGCTGGAGCGCGTCTGCAAGGCTGTGCTCCTTCTCTGCTGCTGC ACTTCTCGTGGCCGTCATCCTCTACTTTGACGTCTACGCCAGCACCTGGCCTTCTT CAGCCGCTTCAGTGCCCGAGGCCCTGCCCATGCCCTCCACCCAGCTGCTAGCAGCAGC AGCAGCAGCAGCAACTGCTCCCGGCCCAACGCCACCGCTCTAGCTCCGGGCTCCCTG AGGTCCCCAGTGCCTGCCCGGTCCCACGGCTCCCACGCTGCCACCCTGTCTTGACAC CTCCCCGCTGGTCTTGTGGGCAGACTGCTGATCGAGTTACCTCACCCATGCCCTG GAGCGGGTGCAGAGGGAGAACCAGGCGTGCTCATGGGCGGCCGATACACATCGCCCG ACTGCACCCAGCCAGACGGTGGCGGTTCATCATCCCTTTAGACACCGGGAACACCA CCTGCGCTACTGGCTCCACTATCTACACCCCATCTTGAGGCGGCAGCGGCTGCGCTAC TGCGTCTATGTCATCAACCAGCATGGTGAGGACACCTTCAACCGGGCCAAGCTGCTTA ACGTGGGCTTCTAGAGGCGCTGAAGGAGGATGCCGCTATGACTGCTTCATCTTCAG CGATGTGGACCTGGTCCCCATGGATGACCGCAACCTATACCGCTGCGGCGACCAACCC CGCCACTTTGCCATTGCCATGGACAAGTTTGGCTTCCGGCTTCCCTATGCTGGCTACT TTGGAGGTGTGTGAGGCTGAGTAAGGCTCAGTTTCTGAGAATCAATGGCTTCCCCAA TGAGTACTGGGGCTGGGGTGGCGAGGATGATGACATCTTCAACCGGATCTCCCTGACT GGGATGAAGATCTACGCCAGACATCCGAATTGGCCGCTACCGCATGATCAAGCAGC ACCGCGACAACGATAACGAACCTAACCCCTCAGAGGTTTACCAAGATTCAAAACACGAA GCTGACCATGAAGCGGACGGCATTGGGTGAGTGCAGTACCAGGTCTTGGAGGTGTCT CGGCAACCACTTTCACCAATATCACAGTGGACATTGGGCGGCCTCCGTGCTGGCCCC CTCGGGGCTGACACTAATGGACAGAGGCTCTCGGTGCCGAAGATTGCTGCCAGAGGA CTGACCACAGCTGGCTGGCAGCTGCTCTGTGGAGGACCTCCAGGACTGAGACTGGGC TCTGTTTCCAAAGGCTTCTCACTAGGCCCTTAGCTATACCTGGAAGTTTCAGAACC ACTTTGGGGGCTCTCCGTGGGCAGGCTCTTCAAGTGTGGCCCTCTTTGGAGTCAACC CTCCTTCCCAGCCCTTCCCTTAGCCAGCCCACTGCTCAGGCTCGGCCAGCC CCTGCACTGCCTCGCAGAGTGGCTGGGCTAGGTCACTCCACCTCTCTGTGCCTCAGT TTCCCCCTTGTAGTCCCTTAGGGCTGGAAGGGTGGGAGGTATGCTAGGGGGCAA GTGTCTTCCAGGGGAATTCTCAGCTTGGGAACCCCTTGGCTCCAGGGGAGGG GAAACCTTTTTTCAATCAACATTGTAGGGGGCAAGCTTGGTGCAGCCCTGCTGAGGA GCGAGCCAGGAGGGGACCAGAGGGGATGCTGTGTCGCTGCCTGGGATCTTGGGGTTG GCCTTGCATGGGAGGCAGGTGGGGCTTGATCAGTAAGTCTGGTTCCCGCTCCCTG TCTGAGAGAGGAGGCAGGANCCAGGGCCGGCTTGTGTTGTACATTGCACAGAACT TGTGTGGGTGCTTTAGTAAAAACGTGAATGG		
	ORF Start: ATG at 50		ORF Stop: TGA at 1169
	SEQ ID NO: 78	373 aa	MW at 42072.7kD
NOV4a, CG103241-01 Protein Sequence	MSRLGGTLERVCKAVLLCLLHFLVAVILYFDVYAQHLAFFSRFSARGPAHALHPAA SSSSSSSNCSRPNATASSSGLPEVPSALPGPTAPTLPPCPDTSPPGLVGRLLIEFTSP MPLERVQRENPVLMGGRYTSPDCTPAQTVAVIIPFRHREHRLRYLWHLHPILRRQR LRYCVYVINQHGEDTFNRKLLNVGFLEALKEDAA YDCFIFSDVDLVPMDDRNLYRCG DQPRHFAIAMDKFGFRLPYAGYFPGVSGLSKAQFLRINGFPNEYWGWDGDDDI FNRI SLTGMKISRDIRIGRYRMIKHDRDNDNEPNPQRF TKIQNTKLTMKRDIGISVRYQVL EVSROPLFTNITVDIGRPPSWPPRG		
	SEQ ID NO: 79	1783 bp	
NOV4b, CG103241-02 DNA Sequence	AGCAGCCGGATGCCCGGGCCCACTGGGCGGGCCAGTGGCCGCTTGCGGGATGAGCAGA CTGCTGGGGGGGACGCTGGAGCGCGTCTGCAAGGCTGTGCTCCTTCTCTGCTGCTGC ACTTCTCGTGGCCGTCATCCTCTACTTTGACGTCTACGCCAGCACCTGGCCTTCTT CAGCCGCTTCAGTGCCCGAGGCCCTGCCCATGCCCTCCACCCAGCTGCTAGCAGCAGC AGCAGCAGCAGCAACTGCTCCCGGCCCAACGCCACCGCTCTAGCTCCGGGCTCCCTG AGGTCCCCAGTGCCTGCCCGGTCCCACGGCTCCCACGCTGCCACCTGCTGACAC CTCCCCGCTGGTCTTGTGGGCAGACTGCTGATCGAGTTACCTCACCCATGCCCTG GAGCGGGTGCAGAGGGAGAACCAGGCGTGCTCATGGGCGGCCGATACACATCGCCCG ACTGCACCCAGCCAGACGGTGGCGGTTCATCATCCCTTTAGACACCGGGAACACCA CCTGCGCTACTGGCTCCACTATCTACACCCCATCTTGAGGCGGCAGCGGCTGCGTAC TGCGTCTATGTCATCAACCAGCATGGTGAGGACACCTTCAACCGGGCCAAGCTGCTTA ACGTGGGCTTCTAGAGGCGCTGAAGGAGGATGCCGCTATGACTGCTTCATCTTCGG CGATGTGGACCTGGTCCCCATGGATGACCGCAACCTATACCGCTGCGGCGACCAACCC CGCCACTTTGCCATTGCCATGGACAAGTTTGGCTTCCGGCTTCCCTATGCTGGCTACT		

	TTGGAGGTGTGTCAGGCCTGAGTAAGGCTCAGTTTCTGAGAATCAATGGCTTCCCCAA TGAGTACTGGGGCTGGGGTGGCGAGGATGATGACATCTTCAACCGGTTTACCAAGATT CAAAACACGAAGCTGACCATGAAGCGGGACGACATTGGGTGAGTGGCGGTACCAGGTCT TGGAGGTGTCTCGGCAACCACTCTTCAACCAATATCACAGTGGACATTGGGCGGCCTCC GTCGTGGCCCCCTCGGGGTGACACTAATGGACAGAGGCTCTCGGTGCCGAAGATTGC CTGCCAGAGGACTGACCACAGCCTGGCTGGCAGCTGCTCTGTGGAGGACCTCCAGGAC TGAGACTGGGGCTCTGTTTTCAGGGTCTTCACTAGGCCCCCTAGCTATACCTGGAAG TTTTCAGAACCCACTTTGGGGGCTCTCCGTGGGCAGGCTCTTCAAGTGTGGCCCTCTT TGGAGTCAACCTCTTCCGACCCCCCTCCCCCTAGCCAGCCCCAGTCACTGTCAGG GTCGGCCAGCCCCGCACTGCCTCGCAGAGTGGCCTGGCTAGGTCACTCCACCTCTC TGTGCCCTCAGTTTCCCCCCTTGAGTCCCCCTTAGGGCTTGAAGGGTGGGAGGTATGT CTAGGGGGCAAGTGTCTCTTCCAGGGGGAATTCTCAGCTCTTGGGAACCCCTTGCTC CCAGGGGAGGGGAACCTTTTTTCATTCAACATTGTAGGGGGCAAGCTTTGGTGCGCC CCTGCTGAGGAGCGAGCCAGGAGGGGACCAGAGGGGATGCTGTGTCGCTGCCTGGGA TCTTGGGGTTGGCCTTTGCATGGGAGGCAGGTGGGGCTTGGATCAGTAAGTCTGGTTC CCGCTCCCTGTCTGAGAGAGGAGGCAGGAACCCAGGGCCGGCTTGTGTTGTACATT GCACAGAACTTGTGTGGGTGCTTTAGTAAAAAACGTGAATGG		
	ORF Start: ATG at 50		ORF Stop: TGA at 1064
	SEQ ID NO: 80	338 aa	MW at 37925.0kD
NOV4b, CG103241-02 Protein Sequence	MSRLLGGLTERVCKAVLLCLLHFLVAVILYFDVYAQHLAFFSRFSARGPAHALHPAA SSSSSSSNCSRPNATASSSGLPEVPSALPGPTAPTLPCCPDSPPGLVGRLLIEFTSP MPLERVQRENPGVLMGGRYTSPDCTPAQTVAVIIPFRHREHHLRYWLHYLHPILRRQR LRYCVYVINQHGEDTFNRAKLLNVGFLEALKEDAAYDCFI FGDVDLVPMDDRNLYRCG DQPRHFAIAMDKFGFRLPYAGYFGGVSGLSKAQFLRINGFPNEYWGWGGEDDDIFNRF TKIQNTKLTMRDDIGSVRYQVLEVSROPLFTNITVDIGRPPSWPPRG		
	SEQ ID NO: 81	1119 bp	
NOV4c, CG103241-03 DNA Sequence	ATGAGCAGACTGCTGGGGGGGACGCTGGAGCGCGTCTGCAAGGCTGTGCTCCTTCTCT GCCTGCTGCACTTCTCGTGGCCGTCATCCTCTACTTTGACGTCTACGCCAGCACCT GGCCTTCTTCAGCCGCTTCAGTGCCCGAGGCCCTGCCATGCCCTCCACCCAGCTGCT AGCAGCAGCAGCAGCAGCAACTGCTCCCGCCCCAACGCCACCGCTCTAGCTCCG GGCTCCCTGAGGTCCCCAGTGCCCTGCCCGTCCCACGGCTCCCACGCTGCCACCCCTG TCTGACTCGCCACCTGGTCTTTGTGGGCAGACTGCTGATCGAGTTACCTCACCCATG CCCCTGGAGCGGGTGACAGGGAGAACCAGCGGTGCTCATGGGCGGCCGATACACAC CGCCCGACTGCACCCAGCCAGACGGTGGCGGTGTCATATCCCTTTAGACACCGGGA ACACCACCTGCGCTACTGGCTCCACTATCTACACCCCATCTTGAAGCGGCAGCGGCTG CGCTACGGCGTCTATGTCATCAACCAGCATGGTGAGGACACCTTCAACCGGGCCAAGC TGCTTAACGTGGGCTTCTAGAGCGCTGAAGGAGGATGCCGCCTATGACTGCTTCAT CTTACGGCATGTGGACCTGGTCCCCATGGATGACCGCAACCTATACCGTGGCGCGAC CAACCCCGCCACTTTGCCATTGCCATGGACAAGTTTGGCTTCCGGCTTCCCTATGCTG GCTACTTTGGAGGTGTGTCAGGCCTGAGTAAGGCTCAGTTTCTGAGAATCAATGGCTT CCCAATGAGTACTGGGGCTGGGGTGGCGAGGATGATGACATCTTCAACCGGATCTCC CTGACTGGGATGAAGATCTCACGCCAGACATCCGAATTGGCCGCTACCGCATGATCA AGCACGACCGGACAAAGCATAACGAACCTAACCTCAGAGGTTTACCAAGATTCAAAA CACGAAGCTGACCATGAAGCGGGACGGCATTGGGTGAGTGCAGTACCAGGTCTTGGAG GTGTCTCGGCAACCACTCTTACCAATATCACAGTGGACATTGGGCGGCCTCCGTCGT GGCCCCCTCGGGGCTGA		
	ORF Start: ATG at 1		ORF Stop: TGA at 1117
	SEQ ID NO: 82	372 aa	MW at 41980.7kD
NOV4c, CG103241-03 Protein Sequence	MSRLLGGLTERVCKAVLLCLLHFLVAVILYFDVYAQHLAFFSRFSARGPAHALHPAA SSSSSSSNCSRPNATASSSGLPEVPSALPGPTAPTLPCCPDSPPGLVGRLLIEFTSPM PLERVHRENPGVLMGGRYTPPDCTPAQTVAVIIPFRHREHHLRYWLHYLHPILRRQRL RYGVYVINQHGEDTFNRAKLLNVGFLEALKEDAAYDCFI FSDVDLVPMDDRNLYRCGD QPRHFAIAMDKFGFRLPYAGYFGGVSGLSKAQFLRINGFPNEYWGWGGEDDDIFNRIS LTGMKISRPDIRIGRYRMIKHDRDKHNEPNPQRF TKIQNTKLTMRDDIGSVRYQVLE VSRQPLFTNITVDIGRPPSWPPRG		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 4B.

<b>Table 4B. Comparison of NOV4a against NOV4b and NOV4c.</b>		
<b>Protein Sequence</b>	<b>NOV4a Residues/ Match Residues</b>	<b>Identities/ Similarities for the Matched Region</b>
NOV4b	1..373	336/373 (90%)
	1..338	336/373 (90%)
NOV4c	1..373	367/373 (98%)
	1..372	367/373 (98%)

Further analysis of the NOV4a protein yielded the following properties shown in Table 4C.

<b>Table 4C. Protein Sequence Properties NOV4a</b>	
<b>PSort analysis:</b>	0.8650 probability located in lysosome (lumen); 0.8200 probability located in outside; 0.2030 probability located in microbody (peroxisome); 0.1000 probability located in endoplasmic reticulum (membrane)
<b>SignalP analysis:</b>	Cleavage site between residues 37 and 38

- 5 A search of the NOV4a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 4D.

<b>Table 4D. Geneseq Results for NOV4a</b>				
<b>Geneseq Identifier</b>	<b>Protein/Organism/Length [Patent #, Date]</b>	<b>NOV4a Residues/ Match Residues</b>	<b>Identities/ Similarities for the Matched Region</b>	<b>Expect Value</b>
AAM93215	Human polypeptide, SEQ ID NO: 2618 - Homo sapiens, 257 aa. [EP1130094-A2, 05-SEP-2001]	117..373 1..257	253/257 (98%) 253/257 (98%)	e-153
AAY17862	Human beta-1,4-galactose transferase - Homo sapiens, 398 aa. [JP11137247-A, 25-MAY-1999]	6..366 16..397	204/384 (53%) 247/384 (64%)	e-109
AAB03647	Beta 1,4 galactose transferase protein sequence - Homo sapiens, 385 aa. [WO200034490-A1, 15-JUN-2000]	6..366 3..384	204/384 (53%) 247/384 (64%)	e-109
AAR28838	HeLa cell galactosyltransferase enzyme - Homo sapiens, 398 aa. [GB2256197-A, 02-DEC-1992]	6..366 16..397	204/384 (53%) 247/384 (64%)	e-109

AAR55706	Galactosyltransferase - Homo sapiens, 398 aa. [WO9412646-A, 09-JUN-1994]	6..366 16..397	204/384 (53%) 247/384 (64%)	e-109
----------	--	-------------------	--------------------------------	-------

In a BLAST search of public sequence databases, the NOV4a protein was found to have homology to the proteins shown in the BLASTP data in Table 4E.

Table 4E. Public BLASTP Results for NOV4a				
Protein Accession Number	Protein/Organism/Length	NOV4a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
O60909	Beta-1,4-galactosyltransferase 2 (EC 2.4.1.-) (Beta-1,4-GalTase 2) (Beta4Gal-T2) (b4Gal-T2) (UDP-galactose:beta-N-acetylglucosamine beta- 1,4-galactosyltransferase 2) (UDP-Gal:beta-GlcNAc beta-1,4-galactosyltransferase 2) [Includes: Lactose synthase A protein (EC 2.4.1.22); N-acetyllactosamine synthase (EC 2.4.1.90) (Nal synthetase); Beta-N-acetylglucosaminyl-glycopeptide beta-1,4- galactosyltransferase (EC 2.4.1.38); Beta-N-acetylglucosaminyl-glycolipid beta-1,4-galactosyltransferase (EC 2.4.1.-)] - Homo sapiens (Human), 372 aa.	1..373 1..372	368/373 (98%) 368/373 (98%)	0.0
Q9Z2Y2	Beta-1,4-galactosyltransferase II - Mus musculus (Mouse), 369 aa.	1..373 1..369	338/373 (90%) 354/373 (94%)	0.0
Q92073	Beta-1,4-galactosyltransferase (EC 2.4.1.38) - Gallus gallus (Chicken), 373 aa.	4..373 5..373	278/378 (73%) 317/378 (83%)	e-164
T46511	hypothetical protein DKFZp586M2424.1 - human, 224 aa (fragment).	150..373 1..224	221/224 (98%) 221/224 (98%)	e-132
CAA01685	GALACTOSYLTRANSFERASE - Homo sapiens (Human), 398 aa.	6..366 16..397	204/384 (53%) 247/384 (64%)	e-108

PFam analysis predicts that the NOV4a protein contains the domains shown in the Table 4F.

Table 4F. Domain Analysis of NOV4a

Pfam Domain	NOV4a Match Region	Identities/ Similarities for the Matched Region	Expect Value
Galactosyl_T_2	97..367	169/330 (51%) 268/330 (81%)	5.5e-190

**Example 5.**

The NOV5 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 5A.

Table 5A. NOV5 Sequence Analysis			
	SEQ ID NO: 83	4215 bp	
NOV5a, CG106249-01 DNA Sequence	CGATGGCATCGGTCAAGGTGGCCGTGAGGGTCCGGCCCATGAATCGCAGGGAAAAGGA CTTGGAGGCAAGTTTATTATTCAGATGGAGAAAAGCAAAACGACAATCACAACTTA AAGATACCAGAAGGAGGCACTGGGGACTCAGGAAGAGAACGGACCAAGACCTTCACCT ATGACTTTTCTTTTATTCTGCTGATACAAAACCTACAGACTACGTTTCACAAGAAAT GGTTTTCAAAACCTCCGCACAGATGTCCTGAATTCTGCATTGAAGTTTATAATGCT TGTGCTTTTGCATATGGGCAAACTGGATCTGGAAAGTCCTACGCTATGATGGGAAATT CTGGAGATTCTGGCTTAATACCTCGGATCTGTGAAGGACTCTCCATTCCGATTAATGA AACCACCAGATCGGATGAAGCTTCTTCCGAACCTGAAGTCAGCTCCTTAAAAATTTAT AACGAACGTGTGAGAGATCTACTTCGGCGGAAGTCATCTAAAACCTTCAATTTGAGAG TCCGTGAGCATCCCAAGAAGGCCCTTATGTTGAGGATTATCCAAACATTTAGTACA GAATTATGGTGACGTAGAAGAAGCTTATGGATGCGGGCAATATCAACCGGACCACCGCA GCGACTGGGATGAACGACGTAGTAGCAGGTCTCATGCCATCTTCACCATCAAGTTCA CTCAGGCTAAATTTGATTCTGAAATGCCATGTGAAACCGTCAGTAAGATCCACTTGGT TGATCTTGCCGGAAGTGAGCGTGAGATGCCACCGGAGCCACCGGGGTAGGCTAAAG GAAGGGGAAATATTAACAAGTCCCTTGTGACTCTGGGGAACGTCATTTCTGCCTTAG CTGATTTATCTCAGGATGCTGCAATACTCTTGCAAAGAAGAAGCAAGTTTTCGTGCC TTACAGGGATTCTGTGTTGACTTGGTTGTTAAAAGATAGCCTTGGAGGAAACTCTAAA ACTATCATGATTGCCACCATTTACCTGCTGATGTCAATTATGGAGAAACCTTAAGTA CTCTTCGCTATGCAATAGAGCCAAAAACATCATCAACAAGCCTACCATTAATGAGGA TGCCAACGTCAAACCTTATCCGTGAGCTGCGAGCTGAAATAGCCAGACTGAAAACGCTG CTTGCTCAAGGGAATCAGATTGCCCTCTTAGACTCCCCACAGCTTAAAGTATGGAGG AAAAACTTCAGCAGAATGAAGCAAGAGTTCAAGAATTGACCAAGGAATGGACAAATAA GTGGAATGAAACCCAAAATATTTGAAAGAACAACCTTAGCCCTCAGGAAAGAAGGG ATTGGAGTTGTTTGGATTCTGAAGTGCCTCATTGATGGGCATCGATGATGACCTTT TGAGTACTGGAATCATCTTATATCATTTAAAGGAAGGTCAGACATACGTTGGTAGAGA CGATGCTTCCACGGAGCAAGATATTGTTCTCATGGCCTTGACTTGGAGAGTGAGCAT TGCATCTTTGAAATATCGGGGGACAGTGACTCTGATACCCCTGAGTGGGTCCCAGT GCTCTGTGAATGGTGTTCAGATCGTGGAGGCCACACATCTAAATCAAGGTGCTGTGAT TCTCTTGGGAAGAACCAATATGTTTCGCTTAAACCATCAAAGGAAGCCGCCAAGCTC AGGGAGAAGAGGAAGAGTGGCCTTCTGTCTCTCTTCAGCTTGTCCATGACCGACCTCT CGAAGTCCCGTGAGAACCTGTCTGCAGTCATGTTGTATAACCCGGGACTTGAGTTTGA GAGGCAACAGCGTGAAGAAGTTGAAAAAATAGAAAGTAAAGGAACCTATTGAGGAA ATGGAGGAAAAGCAGAAATCGGACAAGGCTGAAGTGGAGCGGATGCAGCAGGAGGTGG AGACCCAGCGCAAGGAGACAGAAATCGTGCAGCTCCAGATTGCAAGCAGGAGGAGAG CCTCAAACGCCGAGCTTCCACATCGAGAACAAGCTAAAGGATTTACTTGCGGAGAAG GAAAAATTTGAAGAGGAGAGGCTGAGGGAACAGCAGGAAATCGAGCTGCAGAAGAAGA GACAAGAAGAAGAGACCTTTCTCCGCGTCCAAGAAGAACTCAACGACTCAAAGAAGT CAACAACAACGAGAAGGCTGAGAAGTTTCAAGATATTTCAAGAAGTGGACCACTCCAA AAGGAAAAAGATGAACAGTATGCCAAGCTGAACTGGAAAAAAGAGACTAGAGGAGC AGGAGAAGGAGCAGGTATGCTCGTGGCCCATCTGGAAGAGCAGCTCCGAGAGAAGCA GGAGATGATCCAGCTCCTGCGGCGTGGGAGGTACAGTGGGTGGAAGAGGAGAAGAGG GACCTGGAAGGCATTGCGGAATCCCTCCTGCGGGTGAAGGAGGCTCGTGCCGGAGGGG ATGAAGATGGCGAGGAGTTAGAAAAGGCTCAACTGCGTTTCTTCGAATTCAAGAGAAG		

	GCAGCTTGTCAAGCTAGTGAACCTGGAGAAGGACCTGGTTTCAGCAGAAAGACATCCTG AAAAAGAAGTCCAAGAAGAACAGGAGATCCTAGAGTGTAAAAATGTGAACATGACA AAGAATCTAGATTGTTGGAAAAACATGATGAGAGTGTACAGATGTCACGGAAGTGCC TCAAGATTTTCGAGAAAAATAAGCCAGTGGAGTACAGGCTGCAATATAAAGAACGCCAG CTACAGTACCTCCTGCAGAATCACTTGCCAACTCTGTTGGAAGAAAAGCAGAGAGCAT TTGAAATTTCTGACAGAGGCCCTCTCAGCTTAGACAACACTCTTTATCAAGTAGAAAA GGAAATGGAAGAAAAAGAAGAACAGCTTGACAGTACCAGGCCAATGCAAAACAGCTG CAAAAGCTCCAAGCCACCTTTGAATTCAGTCCCAACATTGCACGTCAGGAGGAAAAAG TGAGGAAAAAGGAAAAGGAGATTTTGGAGTCCAGAGAGAAGCAGCAGAGAGAGCGCT GGAGCGGGCCCTGGCCAGGCTGGAGAGGAGACATTCTGCGCTGCAGAGGCACTCCACC CTGGGCACGGAGATTGAAGAGCAGAGGCAGAACTTGCCAGTGTGAACAGTGGCAGCA GAGAGCAGTCAGGTTCCAGGCTAGCCTGGAGGCTGAGCAGGAAGCACTAGAGATGTA CCATGTAGAAAGGTTAGAATATGAAATCCAGCAGCTGAAACAGAAGATTTATGAGGTC GATGGTGTTCAAAAAGATCATCATGGGACCCTGGAAGGGAAGTGGCTTCTTCCAGCT TGCCAGTCAGTGTGAAAAATCACACCTGGTTCCCTCATGGATGCCAGGAGGATCAA TGCTTACATTGAAGAAGAAGTCCAAAGACGCCCTTCAGGATTTGCATCGTGTGATTAGT GAAGGCTGCAGTACATCTGCAGACAGATGAAGGATAATGAGAACTTCACATGGCA CCATTCAACGTAAACTAAAAATATGAGCTGTGTCTGACCTCCTGTGTCTCTGATGCC AGAGCCTGATGCCGCTGCCGTGCGCTAATCATCCCTTGCTCCAGCAAGATCTGGTTCAG CTTCTCTTGTATGGAAAACAGAAATCCCTGATTTAGTTTTGCCAAATGGAGTTCAGG TGTCTACCAATTCCAGACTACCTTGGTTGACATGATTTACTTTCTTCATGGAATAT GGAAGTCAATGTCCCTTCCCTGGCAGAGTTTCACTTACTGCTCTACACAACAGTGAAA GTCATGGGTGACTCTGGCCATGACCACTGCCAGTCGCTAGTCCCTTCTGAACACCCACA TTGCACTGGTGAAGGAAGACTGTGTTTTTATCCACGCATTCTGATCTCGAAACATACC TCCTCCGGGTGCACAAATTTGATGTGATCAAATGCCATGCTTTAAGTGAATTCAGGTGT GTTGTTGTTCCAGAAAAGAAAAATGTGTCAACAGTAGAACTAGTCTTCTTACAGAAAC TCAAACCTTCAGTGGGTTCAGAAATAGTCCACCTGAGCACCTTCAGGAAGCCCCAAA TGTCAGTTGTTCAACCACCCATTGTATCTTCAAGGCAGTCAGAATGTCGCACCTGAG GTCTGGAACTTACTTTCAATTCTCAAGATGAGGCTCTTGGCTAATCTCACATTTGA CAAGACTCTAAGGAGGAGACTTTTAAAGATGCATACAT		
	ORF Start: ATG at 3	ORF Stop: TAA at 4185	
	SEQ ID NO: 84	1394 aa	MW at 160054.1kD
NOV5a, CG106249-01 Protein Sequence	MASVKVAVRVRPMNRREKDLKAKFI IQMEKSKTTITNLKIPEGGTGDSGRERTKTFTY DFSFYADTKTTDYVSQEMVFKTLRLTDVLNSAFVYNACVFAVGQTGSGKSYAMMGNS GDSGLIPRICEGLSIRINETTRSDEASFRTEVSSLKIYNERVRDLLRRKSSKTFNLRV REHPKEGYPYVEDLSKHLVQNYGDVEELMDAGNINRTAATGMNDVSSRSHAIPTIKFT QAKFDSEMPCEVTISKIHLVDLAGSERADATGATGVRLEKGGNINKSLVTLGNVISALA DLSQDAANTLAKKKQVFPYRDSVLTWLLKDSLGGNSKTIIMATIISPADVNYGETLST LRYANRAKNIINKPTINEDANVKLIRELRAEIALKTLQAQNGQIALLDSPALSMEE KLQQNEARVQELTKWNTKNWNETQNLKEQTLALRKEGIGVLDSELPHLIGIDDDLL STGIILYHLKEGQTYVGRDDASTEQDIVLHGLDLESEHCIFENIGGTVTLIPLSGSQ SVNGVQIVEATHLNQGAIVLLGRTNMFRFNHPKEAAKLREKRKSGLLSSFSLSMTDLS KSRENLSAVMLYNPGLFEFERQQREELKLESKRKLI EEMEEKQKSDKAELERMQQEVE TQRKETEIVQLQIRKQEEESLKRSSFHLENKLDLLAEKEKFEEERLREQQEIQLQKKR QEEETFLLRVQBELQRLKELNNNEKAEPQIFQELDQLQKEKDEQYAKLELEKKRLEEQ EKEQVMLVAHLEEQLEKQEMIQLLRRGEVQWVEEEKRDLEGIRESLLRVKEARAGGD EDGEELEKAQLRFFEFKRRQLVKLVNLEKDLVQQKDILKKEVQEEQEILECLKCEHDK ESRLLEKHDESVDVTEVPQDFEIKPVEYRLQYKERQLQYLQNLHPTLLEEKQRAF EILDRGPLSLDNTLYQVEKEMEEKEEQLAQYQANANQLQKLQATFEFTANARQEEKV RKKEKEILESREKQQREALERALARLERRHSALQRHSTLGTIEEQRQKLASVNSGSR EQSGFQASLEAEQEALMEYHVERLEYEIQLKQKIYEVDGVQKDHGHTLEGKVASSSL PVSAEKSHLVPLMDARRINAYIEEVQRRQLDLHRVISEGCTSDATMKDNEKLHNGT IQRKLYELCRDLLCLVMPEDAAACANHPQLQDLVQLSLDWKTEIPDLVLPNGVQV SSKFQTTLVDMIYFLHGNMEVNVPSLAEVQLLLYTTVKVMGDSGHDQCQSLVLLNTHI ALVKEDCVFYPRIRSRNIPPGAQFDVIKCHALSEFRVVEPEKKNVSTVELVFLQKL KPSVSGSRNSPPEHLQEAAPNVQLFTTPLYLQGSQNVAVEVWKLTFNSQDEALWLISHLT RL		
	SEQ ID NO: 85	4502 bp	

NOV5b,  
CG106249-02  
DNA Sequence

CGGCACGAGGGGGATGAGCGATGGCATCGGTCAAGGTGGCCGTGAGGGTCCGCCCCAT  
GAATCGCAGGGAAAAGGACTTGGAGGCCAAGTTCATTATTCAGATGGAGAAAAGCAAA  
ACGACAATCACAAACTTAAAGATACCAGAAGGAGGCACTGGGGACTCAGGAAGAGAAC  
GGACCAAGACCTTCACCTATGACTTTTCTTTTATTCTGCTGATACAAAAAGCCAGA  
TTACGTTTCACAAGAAATGGTTTTCAAACCCCTCGGCACAGATGTCGTGAAGTCTGCA  
TTTGAAGGTTATAATGCTTGTGTCTTGCATATGGGCAAACTGGATCTGGAAGTCAAT  
ACACTATGATGGGAAATCTGGAGATTCTGGCTTAATACCTCGGATCTGTGAAGGACT  
CTTCAGTCGGATAAATGAAACCACCAGATGGGATGAAGCTTCTTTTCGAACTGAAGTC  
AGCTACTTAGAAATTTATAACGAACGTGTGAGAGATCTACTTCGGCGGAAGTCATCTA  
AAACCTTCAATTTGAGAGTCCGTGAGCATCCCAAAGAAGGCCCTTATGTTGAGGATTT  
ATCCAAACATTTAGTACAGAATTATGGTGACGTAGAAGAACTTATGGATGCGGGCAAT  
ATCAACCGGACCACCGCAGCGACTGGGATGAACGACGTGAGTAGCAGGTCTCATGCCA  
TCTTCACCATCAAGTTCAGTCAAGCTAAATTTGATTCTGAAATGCCATGTGAAACCGT  
CAGTAAGATCCACTTGGTTGATCTTCCCGGAAGTGAGCGTGAGATGCCACCGGAGCC  
ACCGGGGTAGGCTAAAGGAAGGGGAAATATTAACAAGTCCCTTGTGACTCTGGGGA  
ACGTCAATTCTGCCCTAGCTGATTTATCTCAGGATGCTGCAAATACTCTGCAAAGAA  
GAAGCAAGTTTTCTGTCCTTACAGGGATTCTGTGTGACTTGGTTGTTAAAGATAGC  
CTTGGAGGAACTCTAAACTATCATGATTGCCACCATTTACCTGCTGATGTCAATT  
ATGGAGAAACCTAAGTACTCTTCGCTATGCAAATAGAGCCAAAACATCATCAACAA  
GCCTACCATTAATGAGGATGCCAAGCTCAAACCTTATCCGTGAGCTGCGAGCTGAAATA  
GCCAGACTGAAAACGCTGCTTGTCTCAAGGGAATCAGATTGCCCTCTTAGACTCCCCCA  
CAGCTTTAAGTATGGAGGAAAACTTCAGCAGAATGAAGCAAGAGTTCAAGAATTGAC  
CAAGGAATGGACAAATAAGTGGAATGAAACCCAAAATATTTTGAAAGAACAACTCTA  
GCCCTCAGGAAAGAAGGGATTGGAGTTGTTTTGGATTCTGAATGCCCTCATTGATTG  
GCATCGATGATGACCTTTTGAGTACTGGAATCATCTTATATCATTTAAAGGAAGGTCA  
GACATACGTTGGTAGAGACGATGCTTCCACGGAGCAAGATATTGTTCTTCATGGCCTT  
GACTTGGAGAGTGAGCATTGCATCTTTGAAAATATCGGGGGACAGTGACTCTGATAC  
CCCTGAGTGGGTCCAGTGCTCTGTGAATGGTGTTGAGATCGTGAGGCCACACATCT  
AAATCAAGGTGCTGTGATTCTCTTGGGAAGAACCAATATGTTTCGCTTTAACCATCCA  
AAGGAAGCCGCAAGCTCAGGGAGAAGAGGAAGAGTGCCCTCTGTCTCTCTTCAGCT  
TGTCCATGACCGACCTCTCGAAGTCCCGTGAGAACCTGTCTGCAGTCATGTTGTATAA  
CCCCGGACTTGAATTTGAGAGGCAACAGCGTGAAGAACTTGAAAAATTAGAAAGTAAA  
AGGAAACTCATAGAAGAAATGGAGGAAAAGCAGAAATCAGACAAGGCTGAAGTGGAGC  
GGATGCAGCAGGAGGTGGAGACCCAGCGCAAGGAGACAGAAATCGTGACGCTCCAGAT  
TCGCAAGCAGGAGGAGAGCTCAAACGCCGAGCTTCCACATCGAGAACAAGCTAAAG  
GATTTACTTGGGAGAAGGAAAAATTTGAAGAGGAGAGGCTGAGGGAACAGCAGGAAA  
TCGAGCTGCAGAAGAAGAGACAAGAAGAAGAGACCTTTCTCCGCTCCAAGAAGAACT  
CCAACGACTCAAAGAACTCAACAACAACGAGAAGGCTGAGAAGTTTCAGATATTTCAA  
GAAGTGGACAGCTCCAAAAGGAAAAAGATGAACAGTATGCCAAGCTTGAAGTGGAAA  
AAAAGAGACTAGAGGAGCAGGAGAAGGAGCAGGTGATGCTGCGCCCATCTGGAAGA  
GCAGCTCCGAGAGAAGCAGGAGATGATCCAGCTCCTGCGGCGTGGGGAGGTACAGTGG  
GTGGAAGAGGAGAAGAGGGACCTGGAAGGCATTGCGGAATCCCTCCTGCGGGTGAAGG  
AGGCTCGTGCCGAGGGGATGAAGATGGCGAGGAGTTAGAAAAGGCTCAACTGCGTTT  
CTTCGAATTCAGAGAAGGCAGCTTGTCAAGCTAGTGAAGTGGAGAAGGACCTGGTT  
CAGCAGAAAAGACATCCTGAAAAAAGAAGTCCAAGAAGAACAGGAGATCCTAGAGTGT  
TAAATGTGAACATGACAAAGAATCTAGATTGTTGGAAAAACATGATGAGAGTGTAC  
AGATGTCACGGAAGTGCTCAAGATTTTCGAGAAAATAAAGCCAGTGGAGTACAGGCTG  
CAATATAAAGAACGCCAGCTACAGTACCTCCTGCAGAAATCACTTGCCAACTCTGTTGG  
AAGAAAAGCAGAGAGCATTTGAAATCTTGACAGAGGCCCTCTCAGCTTAGACAACAC  
TCTTTATCAAGTAGAAAAGGAAATGGAAGAAAAAGAAGAACAGCTTGACAGTACCAG  
GCCAATGCAAACAGCTGCAAAAGCTCCAAGCCACCTTGAATTAAGTCCCAACATTG  
CACGTCAGGAGGAAAAAGTGAGGAAAAAGGAAAAAGGAGATTTTGGAGTCCAGAGAGAA  
GCAGCAGAGAGAGGCGCTGGAGCGGGCCCTGGCCAGGCTGGAGAGGAGACATTTGCG  
CTGCAGAGGCACTCCACCTTGGGCACGAGATTGAAGAGCAGAGGCAGAACTTGCCA  
GTCTGAACAGTGGCAGCAGAGAGCAGTCAGGGCTCCAGGCTAGCTGGAGGCTGAGCA  
GGAAGCCCTGGAGAAGGACCAGGAGAGGTTAGAATATGAAATCCAGCAGCTGAAACAG  
AAGATTTATGAGGTGATGTTTCAAAAAGATCATCATGGGACCCTGGAAGGGAAGG  
TGGCTTCTTCCAGCTTGGCAGTCAGTGTGAAAAATCACACCTGGTTCCCTCATGGA  
TGCCAGGATCAATGCTTACATTGAAGAAGAAGTCCAAAGACGCCCTCAGGATTTGCAT  
CGTGTGATTAGTGAAGGCTGAGTACATCTGCAGACACGATGAAGGATAATGAGAAAC  
TTCACAATGGCACCATTCAACGTAAACTAAAATATGAGCTGTGTGCTGACCTCCTGTG

	TGTCCTGATGCCAGAGCCTGATGCCGCTGCCTGCGCTAATCATCCCTTGCTCCAGCAA GATCTGGTTTCACTTTCTCTTGATTGGAAAACAGAAATCCCTGATTTAGTTTTGCCAA ATGGAGTTTCAAGTGTATCCAAATCCAGACTACCTTGGTTGACATGATTTACTTTCT TCATGGAAATATGGAAGTCAATGTCCCTTCCCTGGCAGAGTTTCACTTACTGCTCTAC ACAACAGTGAAAGTCATGGGTGACTCTGGCCATGACCAGTGCCAGTCGCTAGTCCTTC TGAACACCCACATTGCACTGGTGAAGGAAGACTGTGTTTTTATCCACGCATTCGATC TCGAAACATACCTCCTCCGGGTGCACAATTTGATGTGATCAAATGCCATGCTTTAAGT GAATTCAGGTGTGTTGTTGTTCCAGAAAAGAAAAATGTGTCAACAGTAGAACTAGTCT TCTTACAGAACTCAAACCTTCAGTGGGTTCCAGAAATAGTCCACCTGAGCACCTTCA GGAAGCCCCAAATGTCCAGTTGTTCAACACCCCATTTGTATCTTCAAGGCAGTCAGAAT GTCGCACCTGAGGTCTGGAACTTACTTTCAATTCTCAAGATGAGGCTCTTTGGCTAA TCTCACATTTGACAAGACTTAAGGAGGAGACTTTTAAAGATGCACTACATGTTTTTT GAGATCATTAATAAAATAAGCATTGTGAAAACAGTCAAGGCAATATGAAATATCTCCGT GTAGCTAATTGAATTGGAAGTGGAAAAATGCAGACCTCTAAAATTGAAAATGTAACCTA TTTTAAATATCTACAATAAAATAAAAACAGCTAATAGCAGAGCCCCAATGAAATATCT TTATCATCACCTTGCTTCTATTTCTTGAAACTCAGGCTTGAAATTTGTGCCTGCTTC ATTATTTGTGAGGTGATTAAAGCATTTCTGATTGTT		
	ORF Start: ATG at 21	ORF Stop: TAA at 4197	
	SEQ ID NO: 86	1392 aa	MW at 159799.8kD
NOV5b, CG106249-02 Protein Sequence	MASVKVAVRVRPMNRREKDLKAFIIQMEKSKTTITNLKIPEGGTGDSGRERTKFTFY DFSFYADTKSPDYVSQEMVFKTLGTDVVKSAFEGYNACVFAYGQTSGSKSYTMMGNS GDSGLIPRICEGLFSRINETTRWDEASFRTEVSYLEIYNERVRDLLRRKSSKTFNLRV REHPKEGFPYVEDLSKHLVQNYGDVEELMDAGNINRTTATGMNDVSSRSHAIFTIKFT QAKFDSEMPCE TVSKIHLVDLAGSERADATGATGVRLEKGNINKSLVTLGNVISALA DLSQDAANTLAKKKQVFVPYRDSVLTWLLKDSLGGNSKTIIMATISPADVNYGETLST LRYANRAKNIINKPTINEDANVKLIRELRAETARLKTLLAQGNQIALDSPTALSMEE KLQQNEARVQELTKEWTKWNQNTQNLKEQTLALRKEGIGVVDSELPHLIGIDDDLL STGIILYHLKEGQTYVGRDDASTEQDIVLHGLDLESEHCI FENIGGTVTLIPLSGSQ SVNGVQIVEATHLNQGA VILLGRNMFRFNHPKEAAKLREKRKSGLLSSFSLSMTDLS KSRENLSAVMLYNPGLEFERQQREELEKLESKRKLIEMEEKQKSDKAELERMQQEVE TQRKETEIVQLQIRKQEESEKRRSFHIEKLDLLAEKEKFEEERLREQQEIQLQKKR QEEETFLRVQEEELQRLKELNNNEKAKEKFQIFQELDQLQKEKDEQYAKLELEKKRLEE EKEQVMLVAHLEELREKQEMIQLLRGEVQWVEEEKRDLEGIRESLRVKEARAGGD EDGEELEKAQLRFFEFKRRQLVKLVNLEKDLVQQKDILKKEVQEEQEI LELCKEHDK ESRLLEKHDESVDVTEVPQDFEIKIPVEYRLQYKERQLQYLLQNLHPTLLEEKQRAF EILDRGPLSLDNTLYQVEKEMEKEEQLAQYQANANQLQKLQATFEFTANIARQEEKV RKKEKEILESREKQOREALERALARLERRHSALQRHSTLGTIEEQKQLASLNSGSR EQSGLQASLEAEQEALEKDQERLEYEIQQLKQKIYEVDGVQKDHGHTLEGKVASSSLP VSAEKSHLVPLMDARINAYIEEVQRRQLDLHRVISEGCTSDATMDNEKLHNGTIQ RKLKYELCRDLLCVLMPEPDAAACANHPQLQDLVQLSLDWKTEIPDLVLPNGVQVSS KFQTTLVDMIYFLHGNMEVNPVSLAEVQLLLYTTVKVMGDSGHDQCSVLVLLNTHIAL VKEDCVFYPIRSRNIPPPGAQFDVIKCHALSEFRVCVVPEKKNVSTVELVFLQKLKP SVGSRNSPPEHLQEAPNVQLFTTPLYLQGSQNVAPVWKLTFNSQDEALWLISHLTRL		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 5B.

Table 5B. Comparison of NOV5a against NOV5b.		
Protein Sequence	NOV5a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV5b	1..1394	1375/1394 (98%)
	1..1392	1379/1394 (98%)

Further analysis of the NOV5a protein yielded the following properties shown in Table 5C.

<b>Table 5C. Protein Sequence Properties NOV5a</b>	
PSort analysis:	0.6086 probability located in mitochondrial matrix space; 0.3127 probability located in mitochondrial inner membrane; 0.3127 probability located in mitochondrial intermembrane space; 0.3127 probability located in mitochondrial outer membrane
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV5a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 5D.

<b>Table 5D. Geneseq Results for NOV5a</b>				
<b>Geneseq Identifier</b>	<b>Protein/Organism/Length [Patent #, Date]</b>	<b>NOV5a Residues/ Match Residues</b>	<b>Identities/ Similarities for the Matched Region</b>	<b>Expect Value</b>
ABB79531	Human kinesin motor protein HsKif16b - Homo sapiens, 1375 aa. [US6399346-B1, 04-JUN- 2002]	1..1394 1..1375	1358/1394 (97%) 1362/1394 (97%)	0.0
AAE22525	Human HsKif16b protein - Homo sapiens, 1375 aa. [US6355471-B1, 12-MAR-2002]	1..1394 1..1375	1358/1394 (97%) 1362/1394 (97%)	0.0
ABB79530	Human kinesin motor protein HsKif16b motor domain - Homo sapiens, 359 aa. [US6399346-B1, 04-JUN-2002]	1..359 1..359	347/359 (96%) 350/359 (96%)	0.0
AAE22526	Human HsKif16b motor domain fragment - Homo sapiens, 359 aa. [US6355471-B1, 12-MAR-2002]	1..359 1..359	347/359 (96%) 350/359 (96%)	0.0
ABB61704	Drosophila melanogaster polypeptide SEQ ID NO 11904 - Drosophila melanogaster, 1174 aa. [WO200171042-A2, 27-SEP-2001]	20..757 1..737	350/776 (45%) 476/776 (61%)	e-161

5 In a BLAST search of public sequence databases, the NOV5a protein was found to have homology to the proteins shown in the BLASTP data in Table 5E.

<b>Table 5E. Public BLASTP Results for NOV5a</b>				
<b>Protein Accession Number</b>	<b>Protein/Organism/Length</b>	<b>NOV5a Residues/ Match</b>	<b>Identities/ Similarities for the Matched</b>	<b>Expect Value</b>

		Residues	Portion	
Q9HCI2	KIAA1590 protein - Homo sapiens (Human), 1238 aa (fragment).	155..1394 1..1238	1233/1240 (99%) 1234/1240 (99%)	0.0
Q9BQM0	DJ971B4.1.2 (KIAA1590 (Novel protein similar to KIF1 type and other kinesin-like proteins) (Isoform 2)) - Homo sapiens (Human), 797 aa (fragment).	596..1394 1..797	791/799 (98%) 792/799 (98%)	0.0
Q9NXN9	CDNA FLJ20135 fis, clone COL06818 - Homo sapiens (Human), 752 aa (fragment).	202..953 1..752	747/752 (99%) 750/752 (99%)	0.0
Q9BQM1	DJ971B4.1.1 (KIAA1590 (Novel protein similar to KIF1 type and other kinesin-like proteins) (Isoform 1)) - Homo sapiens (Human), 722 aa (fragment).	596..1168 1..571	565/573 (98%) 566/573 (98%)	0.0
Q9BQM5	DJ777L9.1 (KIAA1590 (Novel protein similar to KIF1 type and other kinesin-like proteins)) - Homo sapiens (Human), 429 aa (fragment).	37..434 37..429	378/398 (94%) 382/398 (95%)	0.0

PFam analysis predicts that the NOV5a protein contains the domains shown in the Table 5F.

Table 5F. Domain Analysis of NOV5a			
Pfam Domain	NOV5a Match Region	Identities/ Similarities for the Matched Region	Expect Value
kinesin	9..387	187/421 (44%) 301/421 (71%)	3.8e-152
FHA	478..544	21/80 (26%) 45/80 (56%)	0.025

#### Example 6.

5. The NOV6 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 6A.

Table 6A. NOV6 Sequence Analysis			
	SEQ ID NO: 87	858 bp	
NOV6a, CG106824-01	GCCCAGCATGCTCCTCCTTGCTCCCCAGATGCTGAATCTGCTGCTGCTGGCGCTGCCC GTCCTGGCGAGCCGCGCTACGCGGCCCTCCAGCCCCAGGCCAGGCCCTGCAGCGAG TGGGCATCGTCGGGGGTCAGGAGGCCCCAGGAGCAAGTGGCCCTGGCAGGTGAGCCT		

DNA Sequence	GAGAGTCCACGGCCCATACTGGATGCACTTCTGCGGGGCTCCCTCATCCACCCCF TGGGTGCTGACCGCAGCGCACTGCGTGGGACCGGACGCTCAAGGATCTGGCCGCCCTC GGGTGCAACTGCGGGAGCAGCACCTCTACTACCAGGACCAGCTGCTGCCGGTCAGCA GATCATCGTGCACCCACAGTTCTACACCGCCCAGATCGGAGCGGACATCGCCCTGCT GAGCTGGAGGAGCCGGTGAACGTCTCCAGCCACGTCCACACGGTCACCCTGCCCCI CCTCAGAGACCTTCCCCCGGGGATGCCGTGCTGGGTCACTGGCTGGGGCGATGTGC CCCACCGCATTTCTCTGAAGCAGGTGAAGTCCCATAATGGAAAACACATTTG GACGCAAAATACCACCTTGGCGCCTACACGGGAGACGACGTCCGCATCGTCCGTGAC ACATGCTGTGTGCCGGGAACACCCGGAGGGACTCATGCCAGCAGGGCGACTCCGGAG GCCCTTGGTGTGCAAGGTGAATGGCACCTGGCTGCAGGCGGGCGTGGTCAGCTGGGC GAGGGCTGTGCCAGCCCAACCGGCCTGGCATCTACACCGTGTCACTACTACTTTC ACTGGATCCACCCTATGTCCCCAAAAGCCGTGAGTCAGGCCTGG		
	ORF Start: ATG at 8		ORF Stop: TGA at 845
	SEQ ID NO: 88	279 aa	MW at 30877.5kD
NOV6a, CG106824-01 Protein Sequence	MLLLAPQMLNLLLLLALPVLASRAYAAPAPGQALQRVGIVGGQEA PRSKWPQVSLR HGPYWMHFCGGS LIHPQWVLTAAHC VGP DVKDLAALRVQLREQHLYYQDQLLPVSRI VHPQFYTAQIGADIALLELEEPVNVSSHVHTVTLPPASETFPFGMPCWVTGWGDVLP PFPLKQVKVPIMENHICDAKYHLGAYTGDDVRIVRDDMLCAGNTRRDS CQQGDSGGP VCKVNGTWLQAGVVSWE GCAQPNRPGIYTRVTYYLDWIHHYVPKKP		
	SEQ ID NO: 89	828 bp	
NOV6b, CG106824-04 DNA Sequence	ATGCTGAGCCTGCTGCTGCTGGCGCTGCCCGTCTGGCGAGCCCGGCCTACGTGGCC CTGCCCCAGGCCAGGCCCTGCAGCAAACGGGCATTTGTGGGGGCGAGGAGGCCCCCA GAGCAAGTGGCCCTGGCAGGTGAGCCTGAGAGTCCGCGGCCCATACTGGATGCACCTT TGC GGGGGCTCCCTCATCCACCCCACTGGGTGCTAACCGCGGCGCACTGCGTGGAA CGGACATCAAGGATCTGGCCGCCCTCAGGGTGCAACTGCGGGAGCAGCACCTCTACT CCAGGACCAGCTGCTGCCGGTCAGCAGGATCATCGTGCACCCACAGTTCTACATCAT CAGACCGGGGCGGACATCGCCCTGCTGGAGCTGGAGGAGCCCGTGAACATCTCCAGC ACATCCACACGGTCACGCTGCCCTGCCTCGGAGACCTTCCCCCGGGGATGCCGT CTGGGTCACTGGCTGGGGCGACGTGGACAATAATGAGCGCCTCCACCGCCATTTCC CTGAAGCAGGTGAAGGTCCCATAATGGAAAACACATTTGTGACGCAAAATACCAC TTGGCGCCTACACGGGAGACGACGTCCGCATCGTCCGTGACGACATGCTGTGTGCCG GAACACCCGGAGGGACTCATGCCAGGGCGACTCCGGAGGGCCCTGGTGTGCAAGGT AATGGCACCTGGCTGCAGGCGGGCGTGGTCAGCTGGGGCGAGGGCTGTGCCAGGCC ACCGGCCTGGCATCTACACCGGTGTCACTACTACTTGGACTGGATCCACCCTATG CCCCAAAAGCCGTGA		
	ORF Start: ATG at 1		ORF Stop: TGA at 826
	SEQ ID NO: 90	275 aa	MW at 30605.0kD
NOV6b, CG106824-04 Protein Sequence	MLSLLLLALPVLASPAYVAPAPGQALQQTGIVGGQEA PRSKWPQVSLRV RGPYWMHI CGGSLIHPQWVLTAAHCVEPDIKD LAALRVQLREQHLYYQDQLLPVSRIIVHPQFYI QTGADIALLELEEPVNISSHIHTVTLPPASETFPFGMPCWVTGWGDVNNERLPPPF LKQVKVPIMENHICDAKYHLGAYTGDDVRIVRDDMLCAGNTRRDS CQQGDSGGPLVCK NGTWLQAGVVSWE GCAQPNRPGIYTRVTYYLDWIHHYVPKKP		
	SEQ ID NO: 91	828 bp	
NOV6c, CG106824-02 DNA Sequence	ATGCTGAATCTGCTGCTGCTGGCGCTGCCCGTCTGGCGAGCCGCGCTACGCGGCC CTGCCCCAGGCCAGGCCCTGCAGCAGTGGGCATCGTCGGGGGTGAGGAGGCCCCCA GAGCAAGTGGCCCTGGCAGGTGAGCCTGAGAGTCCACGGCCCATACTGGATGCACCT TGCGGGGGCTCCCTCATCCACCCCACTGGGTGCTGACCGCAGCGCACTGCGTGGGA CGGACGTCAAGGATCTGGCCGCCCTCAGGGTGCAACTGCGGGAGCAGCACCTCTACT CCAGGACCAGCTGCTGCCGGTCAGCAGGATCATCGTGCACCCACAGTTCTACACCGC CAGATCGGAGCGGACATCGCCCTGCTGGAGCTGGAGGAGCCGGTGAACGTCTCCAGCC ACGTCCACACGGTCACCTGCCCCCTGCCTCAGAGACCTTCCCCCGGGGATGCCGTG CTGGGTCACTGGCTGGGGCGATGTGGACAATGATGAGCGCCTCCACCGCCATTTCT CTGAAGCAGGTGAAGGTCCCATAATGGAAAACACATTTGTGACGCAAAATACCAC TTGGCGCCTACACGGGAGACGACGTCCGCATCGTCCGTGACGACATGCTGTGTGCCG GAACACCCGAGGGACTCATGCCAGGGCGACTCCGGAGGGCCCTGGTGTGCAAGGT AATGGCACCTGGCTGCAGGCGGGCGTGGTCAGCTGGGGCGAGGGCTGTGCCAGGCCA		

	ACCGGCCTGGCATCTACACCCGTGTCACTACTACTTGGACTGGATCCACCCTATGT CCCCAAAAGCCGTGA		
	ORF Start: ATG at 1		ORF Stop: TGA at 826
	SEQ ID NO: 92	275 aa	MW at 30514.9kD
NOV6c, CG106824-02 Protein Sequence	MLNLLLLLALPVLASRAYAAPAPGQALQRVGIVGGQEAPRSKWPQVSLRVHGPYWMHF CGGSLIHPQWVLTAAHCVGPDVKDLAALRVQLREQHLYYQDQLLPVSRIIVHPQFYTA QIGADIALLELEBPVNVSSHVHTVTLPPASETFPFGMPCWVTGWGDVNDERLPPFPF LKQVKVPIIMENHICDAKYHLGAYTGDDVRIVRDDMLCAGNTRRDSQGDSSGGLVCKV NGTWLQAGVVSWEGCAQPNRPGIYTRVTYYLDWIHHYVPPKP		
	SEQ ID NO: 93	1145 bp	
NOV6d, CG106824-03 DNA Sequence	GGCCAGGATGCTGAATCTGCTGCTGCTGGCGCTGCCGTCCTGGCGAGCCGCGCCTAC GCGGCCCCTGCCCCAGGCCAGGCCCTGCAGCGAGTGGGCATCGTTGGGGGTGAGGAGG CCCCCAGGAGCAAGTGGCCCTGGCAGGTGAGCCTGAGAGTCCACGGCCCCATACTGGAT GCACTTCTGCGGGGGCTCCCTCATCCACCCCCAGTGGGTGCTGACCGCAGCGCACTGC GTGGGACCGGACGCTCAAGGATCTGGCCGCCCTCAGGGTGCACTGCGGGAGCAGCACC TCTACTACCAGGACCAGCTGCTGCCGCTCAGCAGGATCATCGTGACCCACAGTTCTA CACCGCCCAGATCGGAGCGGACATCGCCCTGCTGGAGCTGGAGGAGCCGGTGAAGGTC TCCAGCCACGTCCACACGGTCACCTGCCCCCTGCCTCAGAGACCTTCCCCCGGGGA TGCCGTGCTGGGTCACTGGCTGGGGCGATGTGGACAATGATGAGCGCCTCCCACCGCC ATTTCCTCTGAAGCAGGTGAAGGTCCCCATAATGGAAAACACATTGTGACGCAAAA TACCACCTTGGCGCCTACACGGGAGACGACGTCCGCATCGCTGCAGCAGCATGCTGT GTGCCGGGAACACCCGGAGGGACTCATGCCAGGGCGACTCCGGAGGGCCCTGGTGTG CAAGGTGAATGGCACCTGGCTGCAGGCGGGCGTGGTCAGCTGGGGCGAGGGCTGTGCC CAGCCCAACCGGCTGGCATCTACACCCGTGTCACTACTACTTGGACTGGATCCACC ACTATGTCCCCAAAAGCCGTGAGTCAGGCCTGGGTGGCCACCTGGGTCACTGGAGG ACCAACCCCTGCTGTCTCAAAACACCACTGCTTCTACCCAGGTGGCGACTGCCCCCA CACCTTCCCTGCCCGTCTCTGAGTGCCCTTCTGTCTTAAGCCCCCTGCTCTCTTCT GAGCCCCCTTCCCTGTCTGTGAGGACCTTCCCATCTGAGCCCCCTTCCCTGTCTTA AGCCTGACGCCTGCACCGGGCCCTCCGGCCCTCCCTGCCAGGCAGCTGGTGGTGGG CGCTAATCTCTCTGAGTGTCTGGACCTCATTAAAGTGCATGGAA		
	ORF Start: ATG at 8		ORF Stop: TGA at 833
	SEQ ID NO: 94	275 aa	MW at 30528.9kD
NOV6d, CG106824-03 Protein Sequence	MLNLLLLLALPVLASRAYAAPAPGQALQRVGIVGGQEAPRSKWPQVSLRVHGPYWMHF CGGSLIHPQWVLTAAHCVGPDVKDLAALRVQLREQHLYYQDQLLPVSRIIVHPQFYTA QIGADIALLELEBPVKVSSHVHTVTLPPASETFPFGMPCWVTGWGDVNDERLPPFPF LKQVKVPIIMENHICDAKYHLGAYTGDDVRIVRDDMLCAGNTRRDSQGDSSGGLVCKV NGTWLQAGVVSWEGCAQPNRPGIYTRVTYYLDWIHHYVPPKP		

- 5 Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 6B.

Table 6B. Comparison of NOV6a against NOV6b through NOV6d.		
Protein Sequence	NOV6a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV6b	8..279 1..275	257/277 (92%) 262/277 (93%)
NOV6c	8..279 1..275	270/277 (97%) 270/277 (97%)
NOV6d	8..279 1..275	269/277 (97%) 269/277 (97%)

Further analysis of the NOV6a protein yielded the following properties shown in Table 6C.

<b>Table 6C. Protein Sequence Properties NOV6a</b>	
<b>PSort analysis:</b>	0.8650 probability located in lysosome (lumen); 0.6950 probability located in outside; 0.1333 probability located in microbody (peroxisome); 0.1000 probability located in endoplasmic reticulum (membrane)
<b>SignalP analysis:</b>	Cleavage site between residues 21 and 22

A search of the NOV6a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 6D.

<b>Table 6D. Geneseq Results for NOV6a</b>				
<b>Geneseq Identifier</b>	<b>Protein/Organism/Length [Patent #, Date]</b>	<b>NOV6a Residues/ Match Residues</b>	<b>Identities/ Similarities for the Matched Region</b>	<b>Expect Value</b>
AAW63174	Human mast cell tryptase I polypeptide - Homo sapiens, 273 aa. [WO9833812-A1, 06-AUG-1998]	10..279 1..273	268/275 (97%) 268/275 (97%)	e-161
AAW64238	Human mast cell tryptase I - Homo sapiens, 273 aa. [WO9824886-A1, 11-JUN-1998]	10..279 1..273	268/275 (97%) 268/275 (97%)	e-161
AAW63175	Human mast cell tryptase II/beta polypeptide - Homo sapiens, 274 aa. [WO9833812-A1, 06-AUG-1998]	9..279 1..274	268/276 (97%) 268/276 (97%)	e-161
AAW64240	Human mast cell tryptase II/beta - Homo sapiens, 274 aa. [WO9824886-A1, 11-JUN-1998]	9..279 1..274	268/276 (97%) 268/276 (97%)	e-161
AAE14348	Human protease PRS-13 protein - Homo sapiens, 691 aa. [WO200183775-A2, 08-NOV-2001]	7..279 10..283	263/278 (94%) 264/278 (94%)	e-157

- 5 In a BLAST search of public sequence databases, the NOV6a protein was found to have homology to the proteins shown in the BLASTP data in Table 6E.

<b>Table 6E. Public BLASTP Results for NOV6a</b>				
<b>Protein Accession</b>	<b>Protein/Organism/Length</b>	<b>NOV6a Residues/</b>	<b>Identities/ Similarities for</b>	<b>Expect Value</b>

Number		Match Residues	the Matched Portion	
Q15661	Tryptase beta-1 precursor (EC 3.4.21.59) (Tryptase 1) (Tryptase I) - Homo sapiens (Human), 275 aa.	8..279 1..275	270/277 (97%) 270/277 (97%)	e-162
P20231	Tryptase beta-2 precursor (EC 3.4.21.59) (Tryptase 2) (Tryptase II) - Homo sapiens (Human), 275 aa.	8..279 1..275	269/277 (97%) 269/277 (97%)	e-161
C35863	tryptase (EC 3.4.21.59) III precursor - human, 275 aa.	8..279 1..275	267/277 (96%) 267/277 (96%)	e-159
Q96RZ6	Tryptase I - Homo sapiens (Human), 275 aa.	8..279 1..275	266/277 (96%) 267/277 (96%)	e-159
P15157	Alpha-tryptase precursor (EC 3.4.21.59) (Tryptase 1) - Homo sapiens (Human), 275 aa.	8..279 1..275	252/277 (90%) 258/277 (92%)	e-150

Pfam analysis predicts that the NOV6a protein contains the domains shown in the Table 6F.

Table 6F. Domain Analysis of NOV6a			
Pfam Domain	NOV6a Match Region	Identities/ Similarities for the Matched Region	Expect Value
trypsin	39..271	111/264 (42%) 191/264 (72%)	6.4e-89

### Example 7.

5 The NOV7 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 7A.

Table 7A. NOV7 Sequence Analysis			
	SEQ ID NO: 95	842 bp	
NOV7a, CG114327-01 DNA Sequence	GTGGCCGTCCGAGAGCCGAGAGGTGAGGGTGCCCCCGCCTCACCTGCAGAGGGGCCGT TCCGGGCTCGAACCCGGCACCTTCCGGAAAATGGCGGCTGCCAGGCCAGCCTGGGCC GAGTCTCTCCAGGATCCTCTGTCTGTTCTGTGTGACATGCAGGAGAAGTTCCGCCA CAACATCGCTACTTCCACAGATCGTCTCAGTGGCTGCCCGCATGCTCAAGAACACG ACCCTGGACCTCTAGACCGGGGGCTGCAGGTCCATGTGGTGGTGGACGCCTGCTCCT CACGCAGCCAGGTGGACCGGCTGGTGGCTCTGGCCCGCATGAGACAGAGTGGTGCCTT CCTCTCCACCAGCGAAGGGCTCATTCTGCAGCTTGTGGGCGATGCCGTCCACCCCCAG TTCAAGGAGATCCAGAACTCATCAAGGAGCCCGCCCCAGACAGCGGACTGCTGGGCC TCTTCAAGGCCAGAACTCCCTCCTCCACTGAACCTCCAACCTGCTTGAGGGAAGAC CACCTCCTGTCAACCGGACCTCAGTGAAGCCCGTTCCCCCATCCTGGATCCCAA GAGTGGTGGATCCACCAGGAGTGCCGCCCCCTTGTGGGGGGGGCAGGGTGCTGCCT TCCCATTGACAGCTGCTCCCGGAAATGCAAATGAGACTCCTGGAAGTGGGTGGGAA		

	TTGGCTGAGCCAAGATGGAGGCGGGGCTCGGCCCCGGGCCACTTCACGGGGCGGGAAG GGGAGGGGAAGAAGAGTCTCAGACTGTGGGACACGGACTCGCAGAATAAACATATATG TGGCAAAAAAAAAAAAAAAAAAAAAAAAAA		
	ORF Start: ATG at 89		ORF Stop: TGA at 494
	SEQ ID NO: 96	135 aa	MW at 14765.0kD
NOV7a, CG114327-01 Protein Sequence	MAAARPSLGRVLPGSSVLFCDMQEKFRHNIAYFPQIVSVAARMLKNTTLDLLDRGLQ VHVVDACSSRSQVDRLVALARMRQSGAFLSTSEGLILQLVGDAVHPQFKBIQKLIKE PAPDSGLLGLFQGQNSLLH		
	SEQ ID NO: 97	1091 bp	
NOV7b, CG114327-02 DNA Sequence	GAAACGGTAACCAGCCCTGGGAAGCCCGCAAGAGGCCTCAGCGGTGGCCGTCCGAGCG CCGAGAGGTGAGGGTGCCCCGCCTCACCTGCAGAGGGGCCGTTCGGGGTCCGAACCC GGCACCTTCCGGAAAAATGGCGGCTGCCAGGCCAGCCTGGGCCGAGTCTCCAGGAT CCTCTGTCTGTTCTGTGTGACATGCAGGAGAAGTTCCGCCACAACATCGCCTACTT CCCACAGATCGTCTCAGTGGCTGCCCGCATGCTCAAGGTGGCCCGGCTGCTTGAGGTG CCAGTCATGCTGACGGGAGCAGTACCCACAAGGCCTGGGCCCCACGGTGGCCGAGCTGG GACTGAGGGCCTTCGGCCGCTGGCCAAGACCTGCTTCAGCATGGTGCCTGCCCTGCA GCAGGAGCTGGACAGTCGGCCCCAGCTGCGCTCTGTGCTGCTCTGTGGCATTGAGGCA CAGGCCTGCATCTTGAACACGACCCCTGGACCTCCTAGACCGGGGGCTGCAGGTCCATG TGGTGGTGGACGCCTGCTCCTCAGCGACCCAGGTGGACCGGCTGGTGGCTCTGGCCCG CATGAGACAGAGTGGTGCCTTCTCTCCACCAGCGAAGGGCTCATTCTGCAGCTTGTG GGCGATGCCGTCCACCCCCAGTTC AAGGAGATCCAGAAACTCATCAAGGAGCCCGCCC CAGACAGCGGACTGCTGGGCCTCTTCCAAGGCCAGAACTCCCTCCTCCACTGAACTCC AACCTTGCCTTGAGGGAAGACCACCTCCTGTACCCGGACCTCAGTGGGAAGCCCGTT CCCCCATCCCTGGATCCCAAGAGTGGTGCATCCACCAGGAGTGCCGCCCCCTTGTG GGGGGGGGCAGGGTGTCTGCCTTCCCATTTGGACAGCTGCTCCCGGAAATGCAAATGAGA CTCCTGGAACCTGGGTGGGAATTGGCTGAGCCAAGATGGAGGCGGGGCTCGGCCCCGG GCCACTTCACGGGGCGGGAAGGGGAGGGAAGAAGAGTCTCAGACTGTGGGACACGGA CTCGCAGAATAAACATATATGTGGCTGTGAAAAAAAAAAAAAAAAAAAAA		
	ORF Start: ATG at 132		ORF Stop: TGA at 747
	SEQ ID NO: 98	205 aa	MW at 22336.9kD
NOV7b, CG114327-02 Protein Sequence	MAAARPSLGRVLPGSSVLFCDMQEKFRHNIAYFPQIVSVAARMLKVARLLEVPVMLT EQYPQGLGPTVPELGTGLRPLAKTCFSMPALQQELDSRPQLRSVLLCGIEAQACIL NTTLDLLDRGLQVHVVDACSSRSQVDRLVALARMRQSGAFLSTSEGLILQLVGDAVH PQFKEIQKLIKEPAPDSGLLGLFQGQNSLLH		

- 5 Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 7B.

Table 7B. Comparison of NOV7a against NOV7b.		
Protein Sequence	NOV7a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV7b	35..135 99..205	94/107 (87%) 96/107 (88%)

Further analysis of the NOV7a protein yielded the following properties shown in Table 7C.

Table 7C. Protein Sequence Properties NOV7a	
PSort analysis:	0.5108 probability located in mitochondrial matrix space; 0.4500 probability located in cytoplasm; 0.2553 probability located in lysosome (lumen); 0.2357

	probability located in mitochondrial inner membrane
SignalP analysis:	Cleavage site between residues 24 and 25

A search of the NOV7a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 7D.

Table 7D. Geneseq Results for NOV7a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV7a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAM41577	Human polypeptide SEQ ID NO 6508 - Homo sapiens, 173 aa. [WO200153312-A1, 26-JUL-2001]	1..135 39..173	135/135 (100%) 135/135 (100%)	5e-71
AAM39791	Human polypeptide SEQ ID NO 2936 - Homo sapiens, 135 aa. [WO200153312-A1, 26-JUL-2001]	1..135 1..135	135/135 (100%) 135/135 (100%)	5e-71
AAU23364	Novel human enzyme polypeptide #450 - Homo sapiens, 162 aa. [WO200155301-A2, 02-AUG-2001]	6..133 27..154	122/128 (95%) 123/128 (95%)	5e-63
AAB42186	Human ORFX ORF1950 polypeptide sequence SEQ ID NO:3900 - Homo sapiens, 249 aa. [WO200058473-A2, 05-OCT-2000]	6..135 114..249	99/136 (72%) 105/136 (76%)	1e-44
AAG89278	Human secreted protein, SEQ ID NO: 398 - Homo sapiens, 205 aa. [WO200142451-A2, 14-JUN-2001]	35..135 99..205	94/107 (87%) 96/107 (88%)	3e-44

5. In a BLAST search of public sequence databases, the NOV7a protein was found to have homology to the proteins shown in the BLASTP data in Table 7E.

Table 7E. Public BLASTP Results for NOV7a				
Protein Accession Number	Protein/Organism/Length	NOV7a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q96AB3	Similar to hypothetical protein FLJ23469 - Homo sapiens (Human), 205 aa.	35..135 99..205	94/107 (87%) 96/107 (88%)	8e-44

Q9H5G0	CDNA: FLJ23469 fis, clone HSI11914 - Homo sapiens (Human), 221 aa.	46..135 132..221	89/90 (98%) 90/90 (99%)	1e-43
Q9D8T8	0610042E07Rik protein - Mus musculus (Mouse), 131 aa.	47..134 38..126	69/89 (77%) 78/89 (87%)	8e-31
Q9DCC7	0610042E07Rik protein - Mus musculus (Mouse), 210 aa.	47..134 117..205	69/89 (77%) 78/89 (87%)	8e-31
Q20062	F35G2.2 protein - Caenorhabditis elegans, 199 aa.	48..126 118..196	50/79 (63%) 59/79 (74%)	1e-19

PFam analysis predicts that the NOV7a protein contains the domains shown in the Table 7F.

Table 7F. Domain Analysis of NOV7a			
Pfam Domain	NOV7a Match Region	Identities/ Similarities for the Matched Region	Expect Value
Isochorismatase	13..126	22/213 (10%) 86/213 (40%)	0.61

### Example 8.

5 The NOV8 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 8A.

Table 8A. NOV8 Sequence Analysis			
	SEQ ID NO: 99	1349 bp	
NOV8a, CG119418-01 DNA Sequence	TGCGCCAGGATGGAGTTCGTGAAATGCCTTGGCCACCCGAAGAGTTCTACAACCTGG TGCGCTTCCGGATCGGGGGCAAGCGGAAGGTGATGCCCAAGATGGACCAGGACTCGCT CAGCAGCAGCCTGAAAACCTTGCTACAAGTATCTCAATCAGACCAGTCGCAGTTTCGCA GCTGTTATCCAGGCGCTGGATGGGGAAATGCGCAACGCAGTGTGCATATTTTATCTGG TTCTCCGAGCTCTGGACACACTGGAAGATGACATGACCATCAGTGTGGAAAAGAAGGT CCCGCTGTTACACAACCTTTCACTCTTTCCCTTTACCAACCAGACTGGCGGTTTCATGGAG AGCAAGGAGAAGGATCGCCAGGTGCTGGAGGACTTCCCAACGATCTCCCTTGAGTTTA GAAATCTGGCTGAGAAATACCAAAACAGTGATTGCCGACATTTGCCGAGAAATGGGCAT TGGGATGGCAGAGTTTTTGGATAAGCATGTGACCTCTGAACAGGAGTGGGACAAGTAC TGCCACTATGTTGCTGGGTGGTTCGGAATTGGCCTTTCCCGTCTTTTCTCAGCCTCAG AGTTTGAAGACCCCTTAGTTGGTGAAGATACAGAACGTGCCAACTCTATGGGCCTGTT TCTGCAGAAAACAAACATCATCCGTGACTATCTGGAAGACCAGCAAGGAGGAAGAGAG TTCTGGCCTCAAGAGGTTTGGAGCAGGTATGTTAAGAAGTTAGGGGATTTTGCTAAGC CGGAGAATATTGACTTGGCCGTGCAGTGCCTGAATGAACCTTATAACCAATGCACCTGCA CCACATCCCAGATGTCATCACCTACCTTTTCGAGACTCAGAAACCAGAGTGTGTTTAAC TTCTGCGCTATTCCACAGGTGATGGCCATTGCCACTTTGGCTGCCTGTTATAATAACC AGCAGGTGTTCAAAGGGGCAGTGAAGATTCGGAAGGGCAAGCAGTGACCCTGATGAT GGATGCCACCAATATGCCAGCTGTCAAAGCCATCATATATCAGTATATGGAAGAGATT TATCATAGAATCCCCGACTCAGACCCATCTTCTAGCAAAACAAGGCAGATCATCTCCA CCATCCGGACGCAGAATCTTCCCAACTGTCAGCTGATTTCCTCGAAGCCACTACTCCCC CATCTACCTGTGCTTTGTCTGCTTTTGGCTGCCCTGAGCTGGCAGTACCTGACCACT CTCTCCAGGTAACAGAAGACTATGTTTCAGACTGGAGAACACTGATCCCAAAATTGTC		

	CATAGCTGAAGTCCACCATAAAGTGGATTTACTTTTTTTCTTAAAAAAAAAAAAAAAAAAAA AAAAAAAAAAAAAAAAAAAA		
	ORF Start: ATG at 10		ORF Stop: TGA at 1261
	SEQ ID NO: 100	417 aa	MW at 48114.8kD
NOV8a, CG119418-01 Protein Sequence	MEFVKCLGHPEEFYNLVRFRIGGKRKVMKMDQDSLSSSLKTCYKYLNQTSRSFAAVI QALDGEMRNAVCI FYLVLRALDTLEDDMTISVEKKVPLLHNFHSFLYQPDWRFMESKE KDRQVLEDFPTISLEFRNLAEKYQTVIADICRRMGIGMAEFLDKHVTSEQEWDKYCHY VAGLVGIGLSRLFSASEFEDPLVGEDTERANSMLFLQKTNIRDYLEDQGGREFWP QEVWSRYVKKLGDFAKPENIDLA VQCLNELITNALHHIPDVITYLSRLRNQSVFNFCA IPQVMAIATLAACYNNQOVFKGAVKIRKGQAVTLMMDATNMPAVKAIYQYMEEIYHR IPDSDPSSSKTRQIISTIRTQNLPCQLISRSHYSPIYLSFVMLLAALSWQYLTTLSQ VTEDYVQTGEH		

Further analysis of the NOV8a protein yielded the following properties shown in Table 8B.

Table 8B. Protein Sequence Properties NOV8a	
PSort analysis:	0.4500 probability located in cytoplasm; 0.3719 probability located in microbody (peroxisome); 0.1000 probability located in mitochondrial matrix space; 0.1000 probability located in lysosome (lumen)
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV8a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 8C.

5

Table 8C. Geneseq Results for NOV8a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV8a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAW01739	Human squalene synthetase - Homo sapiens, 417 aa. [US5589372-A, 31-DEC-1996]	1..417 1..417	417/417 (100%) 417/417 (100%)	0.0
AAR52606	Human squalene synthase - Homo sapiens, 417 aa. [GB2272442-A, 18-MAY-1994]	1..417 1..417	416/417 (99%) 416/417 (99%)	0.0
ABB57061	Mouse ischaemic condition related protein sequence SEQ ID NO:118 - Mus musculus, 416 aa. [WO200188188-A2, 22-NOV- 2001]	1..413 1..413	365/413 (88%) 395/413 (95%)	0.0
AAR94574	Squalene synthetase from Nicotiana benthamiana - Nicotiana benthamiana. 411 aa.	7..396 8..401	177/403 (43%) 257/403 (62%)	2e-89

	[WO9609393-A1, 28-MAR-1996]			
AAG32432	Arabidopsis thaliana protein fragment SEQ ID NO: 39123 - Arabidopsis thaliana, 404 aa. [EP1033405-A2, 06-SEP-2000]	7..401 2..401	173/406 (42%) 251/406 (61%)	8e-88

In a BLAST search of public sequence databases, the NOV8a protein was found to have homology to the proteins shown in the BLASTP data in Table 8D.

Table 8D. Public BLASTP Results for NOV8a				
Protein Accession Number	Protein/Organism/Length	NOV8a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
P37268	Farnesyl-diphosphate farnesyltransferase (EC 2.5.1.21) (Squalene synthetase) (SQS) (SS) (FPP:FPP farnesyltransferase) - Homo sapiens (Human), 417 aa.	1..417 1..417	417/417 (100%) 417/417 (100%)	0.0
Q96GT0	Farnesyl-diphosphate farnesyltransferase 1 - Homo sapiens (Human), 417 aa.	1..417 1..417	416/417 (99%) 417/417 (99%)	0.0
I38245	farnesyl-diphosphate farnesyltransferase (EC 2.5.1.21), hepatic - human, 417 aa.	1..417 1..417	416/417 (99%) 416/417 (99%)	0.0
I52090	squalene synthase - human, 417 aa.	1..417 1..417	415/417 (99%) 417/417 (99%)	0.0
P53798	Farnesyl-diphosphate farnesyltransferase (EC 2.5.1.21) (Squalene synthetase) (SQS) (SS) (FPP:FPP farnesyltransferase) - Mus musculus (Mouse), 416 aa.	1..413 1..413	365/413 (88%) 395/413 (95%)	0.0

PFam analysis predicts that the NOV8a protein contains the domains shown in the Table 8E.

Table 8E. Domain Analysis of NOV8a			
Pfam Domain	NOV8a Match Region	Identities/ Similarities for the Matched Region	Expect Value
SQS_PSY	47..334	115/317 (36%) 280/317 (88%)	6.5e-154

**Example 9.**

The NOV9 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 9A.

<b>Table 9A. NOV9 Sequence Analysis</b>			
	SEQ ID NO: 101	2106 bp	
NOV9a, CG120359-01 DNA Sequence	ATGGGGCTTCCTGAGGAGCGGGTCCGGAGCGGCAGCGGGAGCCGGGGCCAGGAGGAAG CTGGAGCCCGAGGCGGGCGCGGAGTTGGTCTCCGCCGCCGAGGTGAGCCGCTCCGC GCACGTCCCCTCGCTGCAGCGCTACCGCGAGCTGCACCGCGCTCCGTGGAGGAGCCG CGGGAATTCCTGGGGAGACATTGCCAAGGAATTTACTGGAAGACTCCATGCCCTGGCC CATTCTTCGGTACAACCTTTGATGTGACTAAAGGGAAAATCTTCATTGAGTGGATGAA AGGAGCAACTACCAACATCTGCTACAATGTACTGGATCGAAATGTCCATGAGAAAAAG CTTGGAGATAAAGTTGCTTTTACTGGGAGGGCAATGAGCCAGGGGAGACCACTCAGA TCACATACCATCAGCTTCTGGTCCAAGTGTGTGCTAGTTCAGCAATGTTCTCCGAAAACA GGGCATTTCAGAAAGGGGACCGAGTGGCCATCTACATGCCTATGATCCCGAGAGCTTGTG GTGGCCATGCTGGCATGTGCCCGCATTTGGGGCTTTCAGTCCATTGTGTTTGCAGGCT TCTCTTCAGAGTCTCTATGTGAACGGATCTTGGATTCCAGCTGCAGTCTTCTCATCAC TACAGATGCTTCTACAGGGGGGAAAAGCTTGTGAACCTGAAGGAGCTGGCTGACGAG GCCCTGCAGAAGTGTGAGGAGAAGGGTTTCCAGTAAGATGCTGCATTGTGGTCAAGC ACCTGGGGCGGGCAGAGCTCGGCATGGGTGACTCCACGAGCAGTCCCCCAATTAA GAGGTCATGCCAGATGTGCAGATCTCATGGAACCAAGGGATTGACTTGTGGTGGCAT GAGCTCATGCAAGAGGCAGGGGATGAGTGTGAGCCGAGTGGTGTGATGCCGAGGACC CACTCTTCATCTGTACACCAAGTGGCTCCACAGGCAAACCAAGGGTGTGGTTACAC AGTTGGGGGCTACATGCTCTATGTAGCCACAACCTTCAAGTATGTGTTTGAATTCAT GCAGAGGATGTGTTCTGGTGCACGGCAGACATTGGTTGGATCACTGGTCATTCTTACG TCACCTATGGGCCCACTGGCCAATGGTGCCACCAAGTGTGTTTGTGAGGGGATTCCAC ATATCCGACGTAACCGCCTGTGGAGCATTGTGGACAAATACAAGGTGACCAAGTTC TACACAGACCCACAGCCATCCGTCTGCTCATGAAGTTTGGAGATGAGCTGTACCA AGCATAGCCGGGCATCCTTGCAGGTGTTAGGCACAGTGGGTGAACCCATCAACCTGA GGCTTGGCTATGGTACCACCGGGTGGTAGGTGCCAGCGCTGCCCATCGTGGACACC TTCTGGCAAAACAGAGACAGGTGGCCACATGTTGACTCCCTTCTGGTGGCCACACCCA TGAAACCCGGTTCTGCTACTTTCCCATTTCTTGGTGTAGCTCCTGCAATCCTGAATGA GTCCGGGGAAGAGTTGGAAGGTGAAGCTGAAGGTTATCTGGTGTTCAGCAGCCCTGG CCAGGGATCATGCGCACAGTCTATGGGAACCAAGACGCTTGGAGACAACCTACTTTA AGAAGTTTCTGGATACTATGTTACAGGAGATGGCTGCCAGCGGGACAGGATGGCTA TTAATGATCACTGGCAGGATTGATGACATGCTCAATGTATCTGGACACCTGCTGAGT ACAGCAGAGGTGGAGTCAGCACTTGTGGAACATGAGGCTGTTGCAGAGGCAGCTGTGG TGGCCACCCCTCATCTGTGAAGGGTGAATGCCTCTACTGCTTGTACCTTGTGTGA TGGCCACACCTTCAGCCCAAGCTCACCGAGGAGCTCAAGAAGCAGATTAGAGAAAAG ATTGGCCCCATTGCCACACCAGACTACATCCAGAATGCACCTGGCTTGCTAAACCC GCTCAGGGAAAATCATGAGGCGAGTGCTTCGGAAGATTGCTCAGAATGACCATGACCT CGGGGACATGTCTACTGTGGCTGACCCATCTGTCATCAGTACCTCTCAGCCACCGC TGCCTGACCATCCAGTGA		
	ORF Start: ATG at 1		ORF Stop: TGA at 2104
	SEQ ID NO: 102	701 aa	MW at 78578.9kD
NOV9a, CG120359-01 Protein Sequence	MGLPEERVRSRGSRGQEEAGAGGRARSWSPPPEVSRSAHVPSLQRYRELHRRSVVEEP REFWDIAKEFYWKTPCPGPFRLRYNFDVTKGKIFIEWMKGATTNICYNVLDNRNVHEKK LGDKVAFYWEGNEPGETTQITYHQLLVQVCQFSNVLKQGIQKGDRAVAYMPMIPELV VAMLACARIGALHSIVFAGFSSESLCERILDSSCSLLITDAFYRGEKLVNLKELADE ALQKCQEKGFVPRCCIVVKHLGRAELMGDSTSQSPPIKRSCPVDQISWNQGIDLWWH ELMQEAGDECEPEWDAEDPLFILIYTSGSTGKPKGVVHTVGGYMLYVATTFKYVDFH AEDVFWCTADIGWITGHSYVTYGPLANGATSVLFEGIPTYPDVNRLWSIVDKYKVTKF YTAPTAIRLLMKFGDEPVTKHSRASLQVLGTVEGEPINPEAWLWYHRVVGARCPVDT FWQETETGGHMLTPLPGATPMKPGSATFPFFGVAPAILNESGEELEGEAEGLVFKQPW PGIMRTVYGNHERFETTYFKKPPGYVYVTDGCGQDQDGYWITGRIDDMNLVSGHLLS TAEVESALVEHEVAEAAVVGHPHPVKGECLYCFVTLCDGHTFSPKLTTELKKQIREK		

	IGPIATPDYIQNAPGLPKTRSGKIMRRVLRKIAQNDHDLGDMSTVADPSVISHLFSHR CLTIQ		
	SEQ ID NO: 103	2125 bp	
NOV9b, 277685717 DNA Sequence	CACCGGATCCACCATGGGGCTTCTGAGGAGCGGGTCCGGAGCGGCAGCGGGAGCCGG GGCCAGGAGGAAGCTGGAGCCGGAGGCCGGCGCGGAGTTGGTCTCCGCCGCCGAGG TCAGCCGCTCCGCGCACGTCCCCTCGTGCAGCGCTACCGCGAGCTGCACCGGCGCTC CGTGGAGGAGCCGCGGAATTCCTGGGGAGACATTGCCAAGGAATTTTACTGGAAGACT CCATGCCCTGGCCATTCTCTCGGTACAACCTTGTATGTGACTAAAGGGAAAATCTTTA TTGAGTGGATGAAAGGAGCAACTACCAACATCTGCTACAATGTACTGGATCGAAATGT CCATGAGAAAAAGCTTGGAGATAAAGTTGCTTTTACTGGGAGGGCAATGAGCCAGGG GAGACCACTCAGATCACATACCATCAGCTTCTGGTCCAAGTGTGTGCTCAGTTCAGCAATG TTCTCCGAAAACAGGGCATTGAGAAGGGGACCGAGTGGCCATCTACATGCCTATGAT CCCAGAGCTTGTGGTGGCCATGCTGGCATGTGCCCGCATTTGGGGCTTTGCACTCCATT GTGTTTGCAGGCTTCTCTTCAAGTCTCTATGTGAACGGATCTTGGATTCCAGCTGCA GTCTTCTCATCACTACAGATGCCTTCTACAGGGGGGAAAAGCTTGTGAACCTGAAGGA GCTGGCTGACGAGGCCCTGCAGAAGTGTGAGGAGAAGGGTTTCCAGTAAGATGCTGC ATTGTGGTCAAGCACCTGGGGCGGGCAGAGCTCGGCATGGGTGACTCCACCAGCCAGT CCCCCCAATTAAGAGGTGATCCCGAGATGTGCAGATCTCATGGAACCAAGGGATTGA CTTGTGGTGGCATGAGCTCATGCAAGAGGCAGGGGATGAGTGTGAGCCCGAGTGGTGT GATGCCGAGGACCACTCTTTCATCCTGTACACAGTGGTCCACAGGCAACCCAAAGG GTGTGTTTACACAGTTGGGGGCTACATGCTCTATGTAGCCACAACCTTCAAGTATGT GTTTGACTTCCATGCAGAGGATGTGTTCTGGTGCACGGCAGACATTGGTTGGATCACT GGTCATTCTTACGTACCTATGGGCCACTGGCCAATGGTGCCACCAGTGTGTTTGTGTTG AGGGGATTCCCATATCCGGACGTGAACCGCTGTGGAGCATTGTGGACAAATACAA GGTGACCAAGTTCTACACAGCACCCACAGCCATCCGTCTGCTCATGAAGTTTGGAGAT GAGCCTGTCAACCAAGCATAGCCGGGCATCCTTGCAGGTGTTAGGCACAGTGGGTGAAC CCATCAACCTGAGGCCTGGCTATGGTACCACCGGGTGGTAGGTGCCAGCGCTGCCC CATCGTGGACACCTTCTGGCAAACAGAGACAGGTGGCCATGTTGACTCCCCTTCTCT GGTGCCACACCCATGAAACCGGTTCTGCTACTTTCCCATTTCTTGGTGTAGCTCCTG CAATCCTGAATGAGTCCGGGAAGAGTTGGAAGGTGAAGCTGAAGGTTATCTGGTGT CAAGCAGCCCTGGCCAGGGATCATGCGCACAGTCTATGGGAACCAACGACGCTTTGAG ACAACCTACTTTAAGAAGTTTCTGGATACTATGTTACAGGAGATGGCTGCCAGCGGG ACCAGGATGGCTATTACTGGATCACTGGCAGGATTGATGACATGCTCAATGTATCTGG ACACCTGCTGAGTACAGCAGAGGTGGAGTCAGCACTTGTGGAACATGAGGCTGTTGCA GAGGCAGCTGTGGTGGGCCACCCTCATCCTGTGAAGGGTGAATGCCCTCACTGCTTTG TCACCTGTGTGATGGCCACACCTTACGCCCAAGCTCACGAGGAGCTCAAGAACGCTT GATTAGAGAAAAGATTGGCCCCATTGCCACACCAGACTACATCCAGAATGCACCTGGC TTGCCATAAACCCGCTCAGGAAAATCATGAGGCGAGTGTCTCGGAAGATTGCTCAGA ATGACCATGACCTCGGGACATGTCTACTGTGGCTGACCCATCTGTCTATCAGTCACCT CTTCAGCCACCGCTGCCTGACCATCCAGCTCGAGGGC		
	ORF Start: at 2	ORF Stop: end of sequence	
	SEQ ID NO: 104	708 aa	MW at 79224.6kD
NOV9b, 277685717 Protein Sequence	TGSTMGLPEERVRSRGSGRQEEAGAGGRARSWSPPEVSRSAHVPSLQRYRELHRRS VEEPREFWGDIAKEFYWKTPCPGPFRLRYNFDVTGKIFIEWMKGATTNICYNVLDNRV HEKKLGDKVAFYWEGNEPGETTQITYHQLLVQCQFSNVLRKQGIQKGRVIAIYMPMI PELVVAMLACARIGALHSIVFAGFSSESLCERILDSSCSLLITDIFYRGEKLVNLKE LADEALQKQKEGFPVRCCIVVKHLGRAELMGDSTSQSPPIKRSCPVDQISWNQID LWWHELMQEAGDECEPEWCDADPLFILIYTSGSTGKPKGVVHTVGGYMLYVATTFKYV FDFHAEDVFWCTADIGWITGHSYVTYGPLANGATSVLFEGIPTYPDVNRLWSIVDKYK VTKFYTAPTALRLMKFGDEPVTKHSRASLQVLGTVGEPINPEAWLWYHRVVGARCP IVDTFWQTETGGHMLTPLGATPMKPGSATFPFFGVAPAILNESGEELEGEAEGLVLF KQPWP GIMRTVYGNHERFETTYFKKFPGYVVTGDGQQRDQDGYWITGRIDMNLNVSG HLLSTA EVESALVEHAEVAEAAVVGHPPVKGECLYCFVTLCDGHTFSPKLTTEELKKQ IREKIGPIATPDYIQNAPGLPKTRSGKIMRRVLRKIAQNDHDLGDMSTVADPSVISHL FSHRCLTIQLEG		
	SEQ ID NO: 105	1408 bp	

NOV9c, 277686882 DNA Sequence	CACCGGATCCACATACCATCAGCTTCTGGTCCAAGTGTGTCAAGTTCAGCAATGTTCTC CGAAAACAGGGCATTGAGAAGGGGGACCGAGTGGCCATCTACATGCCTATGATCCCAG AGCTTGTGGTGGCCATGCTGGCATGTGCCCCGATTGGGGCTTGCAGTCCATTGTGTT TGCAGGCTTCTCTTACAGAGTCTCTATGTGAACGGATCTTGGATTCCAGCTGCAGTCTT CTCATCACTACAGATGCCTTCTACAGGGGGGAAAAGCTTGTGAACCTGAAGGAGCTGG CTGACGAGGCCCTGCAGAAGTGTGACGAGAAGGGTTTCCAGTAAGATGCTGCATTGT GGTCAAGCACCTGGGGCGGGCAGAGCTCGGCATGGGTGACTCCACCAGCCAGTCCCCC CCAATTAAGAGGTGATGCCCAGATGTGCAGATCTCATGGAACCAAGGGATTGACTTGT GGTGGCATGAGCTCATGCAAGAGGCAGGGGATGAGTGTGAGCCCGAGTGGTGTGATGC CGAGGACCCACTCTTCATCCTGTACACCAGTGGCTCCACAGGCAAAACCAAGGGTGTG GTTACACAGTTGGGGGCTACATGCTCTATGTAGCCACAACCTTCAAGTATGTGTTTG ACTTCCATGCGAGGATGTGTTCTGGTGCACGGCAGACATTGGTTGGATCACTGGTCA TTCTTACGTCACCTATGGGGCACTGGCCAATGGTGCCACCAGTGTGTTTGTGAGGGG ATTCCCATATATCCGACGTGAACCGCCTGTGGAGCATTGTGGACAAATACAAGGTGA CCAAGTTCTACACAGCACCCACAGCCATCCGCTCTGCTCATGAAGTTTGGAGATGAGCC TGTCACCAAGCATAGCCGGGCATCCTTGCAGGTGTTAGGCACAGTGGGTGAACCCATC AACCCTGAGGCCTGGCTATGGTACCACCGGGTGGTAGGTGCCGAGCGTCCCCCATCG TGGACACCTTCTGGCAAACAGAGACAGGTGGCCACATGTTGACTCCCTTCTGGTGC CACACCCATGAAACCCGGTTCTGCTACTTTCCCATTTTGGTGTAGCTCCTGCAATC CTGAATGAGTCCGGGGAAGAGTTGGAAGGTGAAGCTGAAGGTATCTGGTGTTCAGC AGCCCTGGCCAGGGATCATGCGCACAGTCTATGGGAACCAAGCAACGCTTTGAGACAAC CTACTTTAAGAAGTTTCTGGATACTATGTTACAGGAGATGGCTGCCAGCGGGACCA GATGGCTATTACTGGATCACTGGCAGGATTGATGACATGCTCAATGTATCTGGACACC TGCTGAGTACAGCAGAGGTGGAGTCAGCACTTGTGGAACATGAGGCTGTGTCAGAGGC AGCTGTGCTCGAGGGC		
	ORF Start: at 2		ORF Stop: end of sequence
	SEQ ID NO: 106	469 aa	MW at 52125.0kD
NOV9c, 277686882 Protein Sequence	TGSTYHQLLVQVCQFSNVLKQGIQKGDRAIYMPMIPELVVAMLACARIGALHSIVF AGFSSESLCERILDSSCSLLITDAFYRGEKLVNLKELADEALQKCQEKGFVPRCCIV VKHLGRAELGMGDSTSQSPPIKRSCPVDQISWNOGIDLWWHELMQEAGDECEPEWCD EDPLFILIYTSGSTGKPKGVVHTVGGYMLYVATTFKYVDFHAEDVFWCTADIGWITGH SYVTYGPLANGATSVLFEGIPTYPDVNRLWSIVDKYKVTKFYAPTAIRLLMKFGDEP VTKHSRASLQVLGTVEGPIPEAWLWYHRVGAQRCPIVDTFWQTETGGHMLTPLPGA TPMKPGSATFPFFGVAPAILNESGEELEGEAEGYLVFKQPWPGIMRTVYGNHERFETT YFKKFPGYVVTGDCQRDQDGYWITGRIDMLNVSGHLLSTAEEVESALVEHEAVAEA AVLEG		
	SEQ ID NO: 107	216 bp	
NOV9d, CG120359-02 DNA Sequence	CACCGGATCCACCATGGGGCTTCTGAGGAGCGGGTCCGGAGCGGCAGCGGAGCCGG GGCCAGGAGGAAGCTGGAGCCGAGGCGGGCGCGGAGTTGGTCTCCGCCGCCGAGG TCAGCCGCTCCGCGCACGTCCCTCGCTGCAGCGCTACCGCGAGCTGCACCGGCGCTC CGTGGAGGAGCCGCGGAATTCTGGGAGACATTGCCAAGGAATTTTACTGGAAGACT CCATGCCCTGGCCATTCTTCCGTACAACCTTGTATGTGACTAAAGGAAAAATCTTCA TTGAGTGGATGAAAGGAGCAACTACCAACATCTGCTACAATGACTGGATCGAAATGT CCATGAGAAAAAGCTTGGAGATAAAGTTGCTTTTACTGGTCCACTTCTGGTAATTCA TCCTACAGATATACTTGCAGGAGGGCAATGAGCCAGGGGAGACCACTCAGATCACAT ACCATCAGCTTCTGGTCCAAGTGTGTGAGTTCAGCAATGTTCTCCGAAAACAGGGCAT TCAGAAGGGGGACCGAGTGGCCATCTACATGCCTATGATCCCAGAGCTTGTGGTGGCC ATGCTGGCATGTGCCCAGTGGGGCTTTGCACTCCATTGTGTTTGCAGGCTTCTCTT CAGAGTCTCTATGTGAACGGATCTTGGATTCCAGCTGCAGTCTTCTCATCTACAGA TGCTTCTACAGGGGGGAAAAGCTTGTGAACCTGAAGGAGCTGGCTGACGAGGCCCTG CAGAAGTGTGAGAGAAGGGTTTCCAGTAAGATGCTGCATTGTGGTCAAGCACCTGG GGCGGCAGAGCTCGGCATGGGTGACTCCACCAGCCAGTCCCCCAATTAAGAGGTC ATGCCAGATGTGCAGATCTCATGGAACCAAGGGATTGACTTGTGGTGGCATGAGCTC ATGCAAGAGGCAGGGGATGAGTGTGAGCCGAGTGGTGTGATGCCGAGGACCACTT TCATCCTGTACACCAGTGGCTCCACAGGCAAAACCAAGGGTGTGGTTACACAGTTGG GGGCTACATGCTCTATGTAGCCACAACCTTCAAGTATGTGTTTGAATCCATGTCAGAG GATGTGTTCTGGTGCACGGCAGACATTGGTTGGATCACTGGTCACTTCTACGTCACCT		

	ATGGGCCACTGGCCAATGGTGCCACCAGTGTGTTTGTGAGGGGATTCACCATATCC GGACGTGAACCGCCTGTGGAGCATTGTGGACAAATACAAGGTGACCAAGTTCTACACA GCACCCACAGCCATCCGTCTGCTCATGAAGTTTGGAGATGAGCCTGTACCAAGCATA GCCGGGCATCCTTGAGGTGTAGGCACAGTGGGTGAACCCATCAACCTTGAGGCCTG GCTATGGTACCACCGGGTGGTAGGTGCCAGCGCTGCCCCATCGTGGACACCTTCTGCG CAAACAGAGACAGGTGGCCACATGTTGACTCCCCCTTCTGGTGCCACCCCATGAAAC CCGGTTCTGCTACTTCCCATTCTTTGGTGTAGCTCCTGCAATCCTGAATGAGTCCGG GGAAGAGTTGGAAGGTGAAGCTGAAGGTTATCTGGTGTTCAGCAGCCCTGGCCAGGG ATCATGCGCACAGTCTATGGGAACCACGAACGCTTTGAGACAACCTACTTTAAGAAGT TTCCTGGATACTATGTTACAGGAGATGGCTGCCAGCGGGACCAGGATGGCTATTACTG GATCACTGGCAGGATTGATGACATGCTCAATGTATCTGGACACCTGCTGAGTACAGCA GAGGTGGAGTCAGCACTTGTGGAACATGAGGCTGTGTCAGAGGCAGCTGTGGTGGGGC ACCTCATCCTGTGAAGGGTGAATGCCTCTACTGCTTTGTACCTTGTGTGATGGCCA CACCTTCAGCCCCAAGCTCACCAGGAGCTCAAGAAGCAGATTAGAGAAAAGATTGGC CCCATTGCCACACCAGACTACATCCAGAATGCACCTGGCTTGCCTAAAACCCGCTCAG GGAAAATCATGAGGCGAGTGCTTCGGAAGATTGCTCAGAATGACCATGACCTCGGGGA CATGTCTACTGTGCTGACCCATCTGTCATCAGTCACCTCTTCAGCCACCGCTGCCTG ACCATCCAGCTCGAGGGC		
	ORF Start: ATG at 14		ORF Stop: at 2156
	SEQ ID NO: 108	714 aa	MW at 80042.4kd
NOV9d, CG120359-02 Protein Sequence	MGLPEERVSRSGSGRQEEAGAGGRARSWSPPEVSRSAHVPSLQRYRELHRRSVVEE REFWGDIAKEFYWKTPCPGPFRLYNFDVTGKGFIIEWMGATTNICYNVLDNRNVEKK LGDKVAFYWSTSGNSSYRYTCREGNPEGETTQITYHQLLVQVCFSNVLRKQGIQKGD RVAIYMPMIPELVVAMLACARIGALHSIVFAGFSSESLCERILDSSCSLLITTDAYR GEKLVNLKELADEALQKCQEKGFVPRCCIIVKHLGRAELGMDSTSSQSPPIKRSCPDV QISWNQGI DLWWHELMQEAGDECEPEWCD AEDPLF ILYTSGSTGKPKGVVHTVGGYML YVATTFKYVDFHAEDVFWCTADIGWITGHSYVTYGPLANGATSVLFEPIPTYPDVNR LWSIVDKYKVKFYTAPTARLLMKFGDEPVTKHSRASLQVLGTVGEPINPEAWLWYH RVVGAQRCPIDTFWQTEGTGHMLTPLPGATPMKPGSATPFFFGVAPAILNESGEELE GEAEGYLVFKQPWPGIMRTVYGNHERFETTYFKKFPGYVVTGDGQORDQDGYWITGR IDDMNLVSGHLLSTA EVESALVEHEAVAEAAVVGHPHPVKGECLYCFVTLCDGHTFSE KLTEELKKQIREKIGPIATPDYIQNAPGLPKTRSGKIMRRVLRKIAQNDHDLGDMSTV ADPSVISHLFSHRCLTIQ		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 9B.

Table 9B. Comparison of NOV9a against NOV9b through NOV9d.		
Protein Sequence	NOV9a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV9b	1..701 5..705	701/701 (100%) 701/701 (100%)
NOV9c	134..600 1..467	464/467 (99%) 465/467 (99%)
NOV9d	1..701 1..714	701/714 (98%) 701/714 (98%)

Further analysis of the NOV9a protein yielded the following properties shown in Table 9C.

Table 9C. Protein Sequence Properties NOV9a
---

PSort analysis:	0.9000 probability located in Golgi body; 0.7900 probability located in plasma membrane; 0.7166 probability located in microbody (peroxisome); 0.2000 probability located in endoplasmic reticulum (membrane)
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV9a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 9D.

Table 9D. Geneseq Results for NOV9a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV9a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAM41491	Human polypeptide SEQ ID NO 6422 - Homo sapiens, 651 aa. [WO200153312-A1, 26-JUL-2001]	59..701 9..651	641/643 (99%) 642/643 (99%)	0.0
AAM39705	Human polypeptide SEQ ID NO 2850 - Homo sapiens, 666 aa. [WO200153312-A1, 26-JUL-2001]	60..701 25..666	641/642 (99%) 641/642 (99%)	0.0
AAB42913	Human ORFX ORF2677 polypeptide sequence SEQ ID NO:5354 - Homo sapiens, 605 aa. [WO200058473-A2, 05-OCT-2000]	96..701 1..605	593/606 (97%) 594/606 (97%)	0.0
AAB94113	Human protein sequence SEQ ID NO:14352 - Homo sapiens, 442 aa. [EP1074617-A2, 07-FEB-2001]	260..701 1..442	441/442 (99%) 442/442 (99%)	0.0
ABB71619	Drosophila melanogaster polypeptide SEQ ID NO 41649 - Drosophila melanogaster, 670 aa. [WO200171042-A2, 27-SEP-2001]	29..696 8..665	420/670 (62%) 522/670 (77%)	0.0

- 5 In a BLAST search of public sequence databases, the NOV9a protein was found to have homology to the proteins shown in the BLASTP data in Table 9E.

Table 9E. Public BLASTP Results for NOV9a				
Protein Accession Number	Protein/Organism/Length	NOV9a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q9NR19	Acetyl-coenzyme A synthetase, cytosolic (EC: 6.2.1.1) (Acetate--	1..701 1..701	701/701 (100%) 701/701 (100%)	0.0

	CoA ligase) (Acyl-activating enzyme) (Acetyl-CoA synthetase) (ACS) (AceCS) - Homo sapiens (Human), 701 aa.			
BAC03849	CDNA FLJ34962 fis, clone NTONG2003897, highly similar to Homo sapiens acetyl-CoA synthetase mRNA - Homo sapiens (Human), 714 aa.	1..701 1..714	699/714 (97%) 700/714 (97%)	0.0
BAC04235	CDNA fis, clone TRACH2001275, highly similar to Mus musculus acetyl-CoA synthetase mRNA - Mus musculus (Mouse), 701 aa.	1..701 1..701	653/701 (93%) 676/701 (96%)	0.0
Q9QXG4	Acetyl-coenzyme A synthetase, cytoplasmic (EC 6.2.1.1) (Acetate--CoA ligase) (Acyl-activating enzyme) (Acetyl-CoA synthetase) (ACS) (AceCS) - Mus musculus (Mouse), 701 aa.	1..701 1..701	651/701 (92%) 673/701 (95%)	0.0
Q96FY7	Unknown (protein for MGC:19474) - Homo sapiens (Human), 442 aa.	260..701 1..442	442/442 (100%) 442/442 (100%)	0.0

PFam analysis predicts that the NOV9a protein contains the domains shown in the Table 9F.

Table 9F. Domain Analysis of NOV9a			
Pfam Domain	NOV9a Match Region	Identities/ Similarities for the Matched Region	Expect Value
AMP-binding	137..599	125/465 (27%) 354/465 (76%)	2.4e-127

#### Example 10.

5 The NOV10 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 10A.

Table 10A. NOV10 Sequence Analysis			
	SEQ ID NO: 109	1958 bp	
NOV10a, CG124907-01 DNA Sequence	GCAGGCCAGCCCCATGCGGAAGCGCAGACGCCGNGCCTGGGCGCTCTGAGATTGTCA CTGCTGTTCCAAGGGCACACGCAGAGGGATTTGGAATTCCTGGAGAGTTGCCTTTGTG AGAAGCTGGAAATATTTCTTTCAATTCCATCTCTTAGTTTTCCATAGGAACATCAAGA AATCATGAACAACTTTGGTAATGAAGAGTTTGACTGCCACTTCCTCGATGAAGGTTT ACTGCCAAGGACATTCTGGACCAGAAAATTAATGAAGTTTCTTCTTCTGATGATAAGG ATGCCTTCTATGTGGCAGACCTGGGAGACATTCTAAGAAACATCTGAGGTGGTTAAA AGCTCTCCCTCGTGTCACCCCTTTTATGCAGTCAAATGTAATGATAGCAAAGCCATC		

	<p>GTGAAGACCCTTGCTGCTACCGGGACAGGATTGACTGTGCTAGCAAGACTGAAATAC  AGTTGGTGCAGAGTCTGGGGGTGCCTCCAGAGAGGATTATCTATGCAAACTCCTTGTA  ACAAGTATCTCAAATTAAGTATGCTGCTAATAATGGAGTCCAGATGATGACTTTTGAT  AGTGAAGTTGAGTTGATGAAAGTTGCCAGAGCACATCCCAAAGCAAAGTTGGTTTTGTC  GGATTGCCACTGATGATTCCAAAGCAGTCTGTCGTCTCAGTGTGAAATTCGGTGCCAC  GCTCAGAACCAGCAGGCTCCTTTTGGAAACGGGCGAAAGAGCTAAATATCGATGTTGTT  GGTGTGAGCTTCCATGTAGGAAGCGGCTGTACCGATCCTGAGACCTTCGTGCAGGCAA  TCTCTGATGCCCCGTGTGTTTTTGACATGGGGGTGAGGTTGGTTTCAGCATGTATCT  GCTTGATATTGGCGGTGGCTTTCCTGGATCTGAGGATGTGAAACTTAAATTTGAAGAG  ATCACCGGCGTAATCAACCCAGCGTTGGACAAATACTTTCGGTCAGACTCTGGAGTGA  GAATCATAGCTGAGCCCGGCAGATACTATGTGTCATCAGCTTTCACGCTTGCAAGTAA  TATCATTGCCAAGAAAATTGTATTAAAGGAACAGACGGGCTCTGATGACGAAGATGAG  TCGAGTGAGCAGACCTTTATGTATTATGTGAATGATGGCGTCTATGGATCATTTAATT  GCATACTCTATGACCACGCACATGTAAAGCCCCCTTCTGAAAAGAGACCTAAACCAGA  TGAGAAGTATTATTCATCCAGCATATGGGGACCAACATGTGATGGCCTCGATCGGATT  GTTGAGCGCTGTGACCTGCCTGAAATGCATGTGGGTGATTGGATGCTCTTTGAAAACA  TGGGCGCTTACACTGTTGCTGCTGCTTACGTTCAATGGCTTCCAGAGGCCGACGAT  CTACTATGTGATGTCAGGGCTGCGTGGCAACTCATGCAGCAATCCAGAACCCCGAC  TTCCACCCGAAGTAGAGGAACAGGATGCCAGCACCTGCGCTGTGTCTTGTGCTGGG  AGAGTGGGATGAAACGCCACAGAGCAGCCTGTGCTTCGGCTAGTATTAAATGTGTAGAT  AGCACTCTGGTAGCTGTTAACTGCAAGTTAGCTTGAATTAAGGGATTGGGGGGACC  ATGTAACCTAATTACTGCTAGTTTTTGAATGTCTTTGTAAGAGTAGGGTCGCCATGAT  GCAGCCATATGGAAGACTAGGATATGGGTCACACTTATCTGTGTTCTATGGAACTA  TTTGAATATTTGTTTATATGGATTTTTATTCACTCTTCAGACACGCTACTCAAGAGT  GCCCTCAGCTGCTGAACAAGCATTGTAGCTTGTACAATGGCAGAAATGGGCCAAAAG  CTTAGTGTTGTGACCTGTTTTTAAATAAAGTATCTTGAAATAACAAAAA  GGGGGGCCGCCCTAGGGGTTCCTCAAGTTTACGTACGCTGCATGG</p>
	<p>ORF Start: ATG at 179</p>
	<p>ORF Stop: TAG at 1562</p>
NOV10a, CG124907-01 Protein Sequence	<p>SEQ ID NO: 110 461 aa MW at 51147.6kD</p> <p>MNNFGNEEFDCFLDEGFTAKDILDKINEVSSDDKDAFYVADLGDILKKHLRLWKA  LPRVTPFFYAVKCNDSKAIKTLAATGTGFDCAASKTEIQLVQSLGVPPERIIYANPCKQ  VSQIKYAANNGVQMMTFDSEVELMKVARAHPKAKLVLRITDDSKAVCRLSVKFGATL  RTSRLLLERAKELNIDVVGVSFHVSGCTDPETFVQAI SDARCVDMDGAEVGF SMYLL  DIGGGFPGSEDEVKLKFEETGVINPALDKYFSPDSGVRIIAEPGRYYVASAFTLAVNI  IAKKIVLKEQTGSDDDESSEQTFMYVNDGVYGSFNCILYDHAHVKPLLQKRPKPE  KYYSSSIWGPTCDGLDRIVERCDLPEMHVGDWMLFENMGAYTVAAASTFNQFORPTIY  YVMSGPAWQLMQQFQNPDPFPEVEEQDASTLPVSCAWESGMKRHRAACASASINV</p>
	<p>SEQ ID NO: 111 1958 bp</p>
NOV10b, CG124907-01 DNA Sequence	<p>GCAGGCCAGCCCCATGGGGAAGCGCAGACGCCGNGCCTGGGCGCTCTGAGATTGTCA  CTGCTGTTCCAAGGGCACACGCAGAGGGATTGGAATTCCTGGAGAGTTGCCTTTGTG  AGAAGCTGGAAATATTTCTTTCAATCCATCTCTTAGTTTTCCATAGGAACATCAAGA  AATCATGAACAACCTTGGTAATGAAGAGTTTGACTGCCACTTCCCTCGATGAAGGTTTT  ACTGCCAAGGACATTCTGGACCAGAAAATTAATGAAGTTTCTTCTTCTGATGATAAGG  ATGCCTTCTATGTGGCAGACCTGGGAGACATTCTAAAGAAACATCTGAGGTGGTTAAA  AGCTCTCCCTCGTGTCAACCCCTTTATGCAGTCAAATGTAATGATAGCAAAGCCATC  GTGAAGACCCTTGCTGCTACCGGGACAGGATTGACTGTGCTAGCAAGACTGAAATAC  AGTTGGTGCAGAGTCTGGGGGTGCCTCCAGAGAGGATTATCTATGCAAACTCCTTGTA  ACAAGTATCTCAAATTAAGTATGCTGCTAATAATGGAGTCCAGATGATGACTTTTGAT  AGTGAAGTTGAGTTGATGAAAGTTGCCAGAGCACATCCCAAAGCAAAGTTGGTTTTGC  GGATTGCCACTGATGATTCCAAAGCAGTCTGTCGTCTCAGTGTGAAATTCGGTGCCAC  GCTCAGAACCAGCAGGCTCCTTTTGGAAACGGGCGAAAGAGCTAAATATCGATGTTGTT  GGTGTGAGCTTCCATGTAGGAAGCGGCTGTACCGATCCTGAGACCTTCGTGCAGGCAA  TCTCTGATGCCCCGTGTGTTTTTGACATGGGGGTGAGGTTGGTTTCAGCATGTATCT  GCTTGATATTGGCGGTGGCTTTCCTGGATCTGAGGATGTGAAACTTAAATTTGAAGAG  ATCACCGGCGTAATCAACCCAGCGTTGGACAAATACTTTCGGTCAGACTCTGGAGTGA  GAATCATAGCTGAGCCCGGCAGATACTATGTGTCATCAGCTTTCACGCTTGCAAGTAA  TATCATTGCCAAGAAAATTGTATTAAAGGAACAGACGGGCTCTGATGACGAAGATGAG  TCGAGTGAGCAGACCTTTATGTATTATGTGAATGATGGCGTCTATGGATCATTTAATT</p>

	GCATACTCTATGACCACGCACATGTAAAGCCCCTTCTGCAAAAGAGACCTAAACCAGA TGAGAAGTATTATTCATCCAGCATATGGGGACCAACATGTGATGGCCTCGATCGGATT GTTGAGCGCTGTGACCTGCCTGAAATGCATGTGGGTGATGGATGCTCTTTGAAAACA TGGGCGCTTACACTGTTGCTGCTGCCTCTACGTTCAATGGCTTCCAGAGGCCGACGAT CTACTATGTGATGTCAGGGCCTGCGTGGCAACTCATGCAGCAATCCAGAACCCCGAC TTCCACCCGAAGTAGAGGAACAGGATGCCAGCACCCCTGCCTGTGCTTGTGCCTGGG AGAGTGGGATGAAACGCCACAGAGCAGCCTGTGCTTCGGCTAGTATTAATGTGTAGAT AGCACTCTGGTAGCTGTTAACTGCAAGTTAGCTTGAATTAAGGGATTGGGGGGACC ATGTAACCTAATTACTGCTAGTTTGAATGTCTTTGTAAGAGTAGGGTCGCCATGAT GCAGCCATATGGAAGACTAGGATATGGGTCACACTTATCTGTGTTCTATGGAACCTA TTTGAATATTTGTTTATATGGATTTTATTCACTCTTCAGACACGCTACTCAAGAGT GCCCCCTCAGCTGCTGAACAAGCATTTGTAGCTTGTACAATGGCAGAATGGGCCAAAAG CTTAGTGTTGTGACCTGTTTAAATAAAGTATCTTGAATAAACAAAAAATAA GGGGGGCCGCCCTAGGGGTTCCTCAAGTTACGTACGCTGCATGG		
	ORF Start: ATG at 179		ORF Stop: TAG at 1562
	SEQ ID NO: 112	461 aa	MW at 51147.6kD
NOV10b, CG124907-01 Protein Sequence	MNNFGNEEFDCFLDEGFTAKDILDQKINEVSSDDKDAFYVADLGDILKKHLRWLKA LPRVTPFYAVKCNDSKAIKTLAATGTGFDCAASKTEIQLVQSLGVPPERIIYANPCKQ VSQIKYAANNGVQMMTFDSEVELMKVARAHPKAKLVLRITATDDSKAVCRLSVKFGATL RTSRLLLERAKELNIDVVGVSFHVSGCTDPETFVQAI SDARC VFDMGAEVGFMYLL DIGGGFPGSEDVKLFEEITGVINPALDKYFSDSGVRI IAEPRGYVVASAFTLAVNI IAKKIVLKEQTGSDDDESSEQTFMYVNDGVYGSFNCILYDHAHVKPLLQKRPKPD KYYSSSIWGPCTDGLDRIVERCDLPEMHVGDWMLFENMGAYTVAAASTFNGFORPTIY YVMSPAWQLMQQFQNPDPPEVEEQDASTLPVSCAWESGMKRHRAACASASINV		
	SEQ ID NO: 113	1416 bp	
NOV10c, 254048022 DNA Sequence	CGCGGATCCACCATGAACAACCTTTGGTAATGAAGAGTTTGACTGCCACTTCTCGATG AAGGTTTTACTGCCAAGGACATTCTGGACCAGAAAATTAATGAAGTTTCTTCTCTGA TGATAAGGATGCCTTCTATGTGGCAGACCTGGGAGACATTCTAAAGAAACATCTGAGG TGGTTAAAGCTCTCCCTCGTGTCAACCCCTTTTATGCAGTCAAATGTAATGATAGCA AAGCCATCGTGAAGACCCTTGCTGCTACCGGGACAGGATTGACTGTGCTAGCAAGAC TGAAATACAGTTGGTGCAGAGCTCGGGGTGCCTCCAGAGAGGATTATCTATGCAAAT CCTTGTAACAAGTATCTCAAATTAAGTATGCTGCTAATAATGGAGTCCAGATGATGA CTTTTGATAGTGAAGTTGAGTTGATGAAAGTTGCCAGAGCACATCCCAAAGCAAAGTT GGTTTTGCGGATTGCCACTGATGATTCAAAGCAGTCTGTCTCAGTGTGAAATTC GGTGCCACGCTCAGAACCAGCAGGCTCCTTTTGAACGGGGCGAAAGAGCTAAATATCG ATGTTGTGTGGTGCAGCTTCCATGTAGGAAGCGGCTGTACCGATCCTGAGACCTTCGT GCAGGCAATCTCTGATGCCCGCTGTGTTTTTGCATGGGGGCTGAGGTTGGTTTCAGC ATGTATCTGCTTGATATTGGCGGTGGCTTTCTGGATCTGAGGATGTGAACTTAAAT TTGAAGAGATCACCGCGTAATCAACCCAGCGTTGGACAAATACTTTCCGTGAGACTC TGGAGTGAGAATCATAGCTGAGCCCGGCAGATACTATGTTGCATCAGCTTTCACGCTT GCAGTTAATATCATTGCCAAGAAAATTGTATTAAGGAACAGACGGGCTCTGATGACG AAGATGAGTCGAGTGAGCAGACCTTTATGTATTATGTGAATGATGGCGTCTATGGATC ATTTAATTGCATACTCTATGACCACGCACATGTAAAGCCCCTTCTGCAAAAGAGACCT AAACCAGATGAGAAGTATTATTCATCCAGCATATGGGGACCAACATGTGATGGCCTCG ATCGGATTGTTGAGCGCTGTGACCTGCCTGAAATGCATGTGGGTGATTGGATGCTCTT TGAAACATGGGCGCTTACACTGTTGCTGCTGCCTCTACGTTCAATGGCTTCCAGAGG CCGACGATCTACTATGTGATGTCAGGGCCTGCGTGGCAACTCATGCAGCAATTCAGA ACCCTGACTTCCACCCGAAGTAGAGGAACAGGATGCCAGCACCCCTGCCTGTGCTTGT TGCCTGGGAGAGTGGGATGAAACGCCACAGAGCAGCCTGTGCTTCGGCTAGTATTAAT GTGTAGGCGGCCGCTTTTTTCTT		
	ORF Start: at 1		ORF Stop: TAG at 1396
	SEQ ID NO: 114	465 aa	MW at 51549.0kD
NOV10c, 254048022 Protein Sequence	RGSTMNFGNEEFDCFLDEGFTAKDILDQKINEVSSDDKDAFYVADLGDILKKHLR WLKALPRVTPFYAVKCNDSKAIKTLAATGTGFDCAASKTEIQLVQSLGVPPERIIYAN PCKQVSQIKYAANNGVQMMTFDSEVELMKVARAHPKAKLVLRITATDDSKAVCRLSVKF GATLRTSRLLLERAKELNIDVVGVSFHVSGCTDPETFVQAI SDARC VFDMGAEVGFS MYLLDIGGGFPGSEDVKLFEEITGVINPALDKYFSDSGVRI IAEPRGYVVASAFTL		

	AVNIIAKKIVLKEQTGSDDDEDESSEQTFMYVNDGVYGSFNCILYDHAHVKPLLQKRP KPDEKYYSSSIWGPTCDGLDRIVERCDLPEMHVGDWMLFENMGAYTVAAASTFNGFQR PTIYYVMSGPAWQLMQQFQNPDPFPEVEEQDASTLPVSCAWESGMKRHRAACASASIN V		
	SEQ ID NO: 115	1410 bp	
NOV10d, 258252457 DNA Sequence	ACCATGGGCCACCATCACCACCATCACAACAACCTTTGGTAATGAAGAGTTTGACTGCC ACTTCCTCGATGAAGGTTTTACTGCCAAGGACATTCTGGACCAGAAAATTAATGAAGT TTCTTCTTCTGATGATAAGGATGCCTTCTATGTGGCAGACCTGGGAGACATTCTAAAG AAACATCTGAGGTGGTTAAAAGCTCTCCCTCGTGTCAACCCCTTTATGCAGTCAAAT GTAATGATAGCAAAGCCATCGTGAAGACCCCTTGCTGCTACCGGGACAGGATTTGACTG TGCTAGCAAGACTGAAATACAGTTGGTGCAGAGTCTGGGGGTGCCTCCAGAGAGGATT ATCTATGCAAATCCTTGTAAACAAGTATCTCAAATTAAGTATGCTGCTAATAATGGAG TCCAGATGATGACTTTTGATAGTGAAGTTGAGTTGATGAAAGTTGCCAGAGCACATCC CAAAGCAAAGTTGGTTTTGCGGATTGCCACTGATGATTCAAAGCAGTCTGTGCTCTC AGTGTGAAATTCGGTGCCACGCTCAGAACCAGCAGGCTCCTTTTGGAACGGGCGAAAG AGCTAAATATCGATGTTGTTGGTGTGAGCTTCCATGTAGGAAGCGGCTGTACCGATCC TGAGACCTTCGTGCAGGCAATCTCTGATGCCCCGTGTGTTTTTGACATGGGGGCTGAG GTTGGTTTTCAGCATGTATCTGCTTGATATTGGCGGTGGCTTTCTCTGGATCTGAGGATG TGAAACTTAAATTTGAAGAGATCACCGGCGTAATCAACCCAGCGTTGGACAAATACTT TCCGTGAGACTCTGGAGTGAGAATCATAGCTGAGCCGGCAGATACTATGTTGCATCA GCTTTCACGCTTGCAAGTTAATATCATTGCCAAGAAAATTGTATTAAAGGAACAGACGG GCTCTGATGACGAAGATGAGTCGAGTGAGCAGACCTTTATGTATTATGTGAATGATGG CGTCTATGGATCATTTAATTGCATACTCTATGACCACGCACATGTAAAGCCCTTCTG CAAAAGAGACCTAAACCAGATGAGAAGTATTATTCATCCAGCATATGGGGACCAACAT GTGATGGCCTCGATCGGATTGTTGAGCGCTGTGACCTGCCTGAAATGCATGTGGGTGA TTGGATGCTCTTTGAAAACATGGGCGCTTACACTGTTGCTGCTGCCTCTACGTTCAAT GGCTTCCAGAGGCCGACGATCTACTATGTGATGTGAGGGCCTGCGTGGCAACTCATGC AGCAATTCAGAACCCCTGACTTCCCACCCGAAGTAGAGGAACAGGATGCCAGCACCTT GCCTGTGCTTGTGCTGGGAGAGTGGGATGAAACGCCACAGAGCAGCCTGTGCTTCG GCTAGTATTAATGTGTAG		
	ORF Start: at 1		ORF Stop: TAG at 1408
	SEQ ID NO: 116	469 aa	MW at 52128.6kD
NOV10d, 258252457 Protein Sequence	TMGHHHHHHNFGNEEFDFCHFLDEGFTAKDILDQKINEVSSSDDKDAFYVADLGDILK KHLRWLKLPRVTPFYAVKCNDSKAI VKTLAATGTGFDCASKTEIQLVQSLGVPPERI IYANPCKQVSQIKYAANNGVQMMTFDSEVELMKVARHPKAKLVLRATDDSKAVCRL SVKFGATLRTSRLLLERAKELNIDVVGVSFHVSGCTDPETFFVQATSDARCVPDMGAE VGFSMYLLDIGGGFPGSEDEVKLKFEEITGVINPALDKYFPSDSGVRI IAEPRYYVAS AFTLAVNIIAKKIVLKEQTGSDDDEDESSEQTFMYVNDGVYGSFNCILYDHAHVKPLL QKRPKPDEKYYSSSIWGPTCDGLDRIVERCDLPEMHVGDWMLFENMGAYTVAAASTFN GFQRPTIYYVMSGPAWQLMQQFQNPDPFPEVEEQDASTLPVSCAWESGMKRHRAACAS ASINV		
	SEQ ID NO: 117	1407 bp	
NOV10e, 258280014 DNA Sequence	ACCATGAACAACCTTTGGTAATGAAGAGTTTGACTGCCACTTCCTCGATGAAGGTTTGA CTGCCAAGGACATTCTGGACCAGAAAATTAATGAAGTTTCTTCTCTGATGATAAGGA TGCTTCTATGTGGCAGACCTGGGAGACATTCTAAAGAAAACATCTGAGGTGGTTAAAA GCTCTCCCTCGTGTCAACCCCTTTTATGCAGTCAAATGTAATGATAGCAAAGCCATCG TGAAGACCCCTTGCTGCTACCGGGACAGGATTTGACTGTGCTAGCAAGACTGAAATACA GTTGGTGCAGAGTCTGGGGGTGCCTCCAGAGAGGATTATCTATGCAAATCCTTGTA CAAGTATCTCAAATTAAGTATGCTGCTAATAATGGAGTCCAGATGATGACTTTTGATA GTGAAGTTGAGTTGATGAAAGTTGCCAGAGCACATCCCAAAGCAAAGTTGGTTTTGCG GATTGCCACTGATGATTCCAAAGCAGTCTGTCTGCTCAGTGTGAAATTCGGTGCCACG CTCAGAACCAGCAGGCTCCTTTTGAACGGGCGAAAGAGCTAAATATCGATGTTGTTG GTGTCAGCTTCCATGTAGGAAGCGGCTGTACCGATCCTGAGACCTTCGTGCAGGCAAT CTCTGATGCCCGCTGTGTTTTGACATGGGGGCTGAGGTTGGTTTCAGCATGTATCTG CTTGATATTGGCGGTGGCTTTCCTGGATCTGAGGATGTGAAACTTAAATTTGAAGAGA TCACCGGCGTAATCAACCCAGCGTTGGACAAATACTTCCGTGAGCTCTGAGTGAG AATCATAGCTGAGCCCGGCAGATACTATGTTGCATCAGCTTTCACGCTTGCAAGTTAAT		

	ATCATTGCCAAGAAAATTGTATTAAAGGAACAGACGGGCTCTGATGACGAAGATGAGT CGAGTGAGCAGACCTTTATGTATTATGTGAATGATGGCGTCTATGGATCATTTAATTG CATACTCTATGACCACGCACATGTAAAGCCCCCTTCTGCAAAAGAGACCTAAACCAGAT GAGAAGTATTATTCATCCAGCATATGGGGACCAACATGTGATGGCCTCGATCGGATTG TTGAGCGCTGTGACCTGCCTGAAATGCATGTGGGTGATTGGATGCTCTTTGAAAACAT GGGCGCTTACACTGTTGCTGCTGCCTCTACGTTCAATGGCTTCCAGAGGCCGACGATC TACTATGTGATGTGAGGCTGCTGGCAACTCATGCAGCAATTCCAGAACCCTGACT TCCCACCCGAAGTAGAGGAACAGGATGCCAGCACCTGCCTGTGTCTTGTGCTGGGA GAGTGGGATGAAACGCCACAGAGCAGCCTGTGCTTCGGCTAGTATTAATGTGCACCAT CACCACCATCACTGA		
	ORF Start: at 1		ORF Stop: TGA at 1405
	SEQ ID NO: 118	468 aa	MW at 52071.6kD
NOV10e, 258280014 Protein Sequence	TMNNFGNEEFDCHFLDEGFTAKDILDQKINEVSSDDKDAFYVADLGDILKKHLRWL ALPRVTPFYAVKCNDSKAIVKTLAATGTGFDCASTEIQLVQSLGVPPERIIYANPCK QVSQIKYAANNGVQMMTFDSEVELMKVARAHPKAKLVLRATDDSKAVCRLSVKFGAT LRTSRLLLERAKELNIDVVGVSFHVSGCTDPETFVQAISDARCVFDMGAEVGFSMYL LDIGGGFPGSEDEVKLKFEIITGVINPALDKYFSDSGVRIIAEPGRYYVASAFTLAVN IIAKKIVLKEQTGSDDDEDESSEQTFMYVNDGVYGSFNCILYDHAHVKPLLQKRPKPD EKYYSSSIWGPTCDGLDRIVERCDLPEMHVGDWMLFENMGAYTVAAASTFNGFQRP TIYYVMSGPAWQLMQQFQNPDPFPEVEEQDASTLPVSCAWESGMKRHRACASASIN VHHHHH		
	SEQ ID NO: 119	1434 bp	
NOV10f, 258330318 DNA Sequence	CACCATCACCACCATCAACAACCTTTGGTAATGAAGAGTTGACTGCCACTTCCTCG ATGAAGGTTTTACTGCCAAGGACATTCTGGACCAGAAAATTAATGAAGTTTCTTCTTC TGATGATAAGGATGCCTTCTATGTGGCAGACCTGGGAGACATTCTAAAGAAACATCTG AGGTGGTTAAAGCTCTCCCTCGTGTCAACCCCTTTTATGCAGTCAAATGTAATGATA GCAAAGCCATCGTGAAGACCCTTGCTGCTACCGGGACAGGATTTGACTGTGCTAGCAA GACTGAAATACAGTTGGTGCAGAGTCTGGGGTGCCTCCAGAGAGGATTATCTATGCA AATCCTTGTAACAAGTATCTCAAATTAAGTATGCTGCTAATAATGGAGTCCAGATGA TGACTTTTGATAGTGAAGTTGAGTTGATGAAAGTTGCCAGAGCACATCCCAAGCAAA GTTGGTTTTCGGATTGCCACTGATGATTCCAAAGCAGTCTGCTGCTCAGTGTGAAA TTCGGTGCCACGCTCAGAACCAGCAGGCTCCTTTTGAACGGGCGAAAGAGCTAAATA TCGATGTTGTTGGTGTGAGCTTCCATGTAGGAAGCGGCTGTACCGATCCTGAGACCTT CGTGCAGGCAATCTCTGATGCCCCTGTGTTTTGACATGGGGGCTGAGGTTGGTTTC AGCATGTATCTGCTTGATATTGGCGGTGGCTTTCCTGGATCTGAGGATGTGAACTTA AATTTGAAGAGATCACCAGCGTAATCAACCAGCGTTGGACAAATACCTTCGTCAGA CTCTGGAGTGAGAATCATAGCTGAGCCCGGCAGATACTATGTCATCAGCTTTCACG CTTGCAAGTTAATATCATTGCCAAGAAAATTGTATTAAAGGAACAGACGGGCTCTGATG ACGAAGATGAGTCGAGTGAGCAGACCTTTATGTATTATGTGAATGATGGCTCTATGG ATCATTTAATTGCATACTCTATGACCACGCACATGTAAAGCCCCCTTCTGCAAAAGAGA CCTAAACCAGATGAGAAGTATTATTCATCCAGCATATGGGGACCAACATGTGATGGCC TCGATCGGATTGTTGAGCGCTGTGACCTGCCTGAAATGCATGTGGGTGATTGGATGCT CTTTGAAAACATGGGCGCTTACACTGTTGCTGCTGCCTCTACGTTCAATGGCTTCCAG AGGCCGACGATCTACTATGTGATGTGAGGCTGCGTGGCAACTCATGCAGCAATTCC AGAACCCTGACTTCCACCCGAAGTAGAGGAACAGGATGCCAGACCCCTGCTGTGTC TTGTGCTGGGAGAGTGGGATGAAACGCCACAGAGCAGCCTGTGCTTCGGCTAGTATT AATGTGTAGGCGGCGCACTCGAGCACCACCACCACCACC		
	ORF Start: at 1		ORF Stop: TAG at 1399
	SEQ ID NO: 120	466 aa	MW at 51839.3kD
NOV10f, 258330318 Protein Sequence	HHHHHHNFGNEEFDCHFLDEGFTAKDILDQKINEVSSDDKDAFYVADLGDILKKHL RWLALPRVTPFYAVKCNDSKAIVKTLAATGTGFDCASTEIQLVQSLGVPPERIIYA NPCKQVSQIKYAANNGVQMMTFDSEVELMKVARAHPKAKLVLRATDDSKAVCRLSVK FGATLRTSRLLLERAKELNIDVVGVSFHVSGCTDPETFVQAISDARCVFDMGAEVG SMYLLDIGGGFPGSEDEVKLKFEIITGVINPALDKYFSDSGVRIIAEPGRYYVASAFT LAVNIIAKKIVLKEQTGSDDDEDESSEQTFMYVNDGVYGSFNCILYDHAHVKPLLQKR PKPDEKYYSSSIWGPTCDGLDRIVERCDLPEMHVGDWMLFENMGAYTVAAASTFNGFQ RPTIIYYVMSGPAWQLMQQFQNPDPFPEVEEQDASTLPVSCAWESGMKRHRACASASI		

	NV		
	SEQ ID NO: 121	1305 bp	
NOV10g, 258330346 DNA Sequence	ACATCATCACCACCATCAAACAACCTTTGGTAATGAAGAGTTTGACTGCCACTTCCTCG ATGAAGGTTTACTGCCAAGGACATTCTGGACCAGAAAATTAATGAAGTTTCTTCTTC TGATGATAAGGATGCCTTCTATGTGGCAGACCTGGGAGACATTCTAAAGAAACATCTG AGGTGGTTAAAGCTCTCCCTCGTGTCACCCCTTTTATGCAGTCAAATGTAATGATA GCAAAGCCATCGTGAAGACCTTGCTGCTACCGGGACAGGATTGACTGTGCTAGCAA GACTGAAATACAGTTGGTGCAGAGTCTGGGGGTGCCTCCAGAGAGGATTATCTATGCA AATCCTTGTAACAAGTATCTCAAATTAAGTATGCTGCTAATAATGGAGTCCAGATGA TGACTTTTGATAGTGAAGTTGAGTTGATGAAAGTTGCCAGAGCACATCCCAAAGCAAA GTTGGTTTTCGGATTGCCACTGATGATTCAAAGCAGTCTGTCGTCTCAGTGTGAAA TTCGGTGCCACGCTCAGAACAGCAGGCTCCTTTTGAACGGGCGAAAGAGCTAAATA TCGATGTTGTTGGTGTCTAGCTTCCATGTAGGAAGCGGCTGTACCGATCCTGAGACCTT CGTGCAGGCAATCTCTGATGCCCGCTGTGTTTTGACATGGGGGCTGAGGTTGGTTTC AGCATGTATCTGCTTGATATTGGCGGTGGCTTCTCGGATCTGAGGATGTGAACTTA AATTTGAAGAGATCACCAGCGTAATCAACCAGCGTTGGACAAATACCTTCGTGAGA CTCTGGAGTGAGAATCATAGCTGAGCCCGGCAGATACTATGTTGCATCAGCTTTCACG CTTGCACTTAATATCATGTGCCAAGAAAATTGTATTAAAGGAACAGACGGGCTCTGATG ACGAAGATGAGTCGAGTGAGCAGACCTTTATGTATTATGTGAATGATGGCGTCTATGG ATCATTTAATTGCATACTCTATGACCACGCACATGTAAAGCCCTTCTGCAAAAGAGA CCTAAACCAGATGAGAAGTATTATTCATCCAGCATATGGGGACCAACATGTGATGGCC TCGATCGGATTGTTGAGCGCTGTGACCTGCCTGAAATGCATGTGGGTGATTGGATGCT CTTTGAAACATGGGCGCTTACACTGTTGCTGCTGCCTCTACGTTCAATGGCTTCCAG AGGCCGACGATCTACTATGTGATGTCAGGGCCTGCGTGGCAACTCATGCAGTAGGCGG CCGCACTCGAGCACCACCACCACCAC		
	ORF Start: at 1		ORF Stop: TAG at 1270
	SEQ ID NO: 122	423 aa	MW at 46885.9kD
NOV10g, 258330346 Protein Sequence	TSSPPSNFNGNEEFDCFLDEGFTAKDILDQKINEVSSSDDKDAFYVADLGDILKKHL RWLKALPRVTPFYAVKCNDSKAI VKTLAATGTGFDCAKTEIQLVQSLGVPPERIIYA NPCKQVSIKYAANNVQMMTFDSEVELMKVARAHPKAKLVLR IATDDSKAVCRLSVK FGATLRSRLLLERAKELNIDVGVSFHVGSGCTDPETFVQAI SDARCVFDMGAEVGF SMYLLDIGGGFPGSEDVKLKFEEITGVINPALDKYFSDSGVRIIAEPGRYYVASAFT LAVNIIAKKIVLKEQTGSDDDEDESSEQTFMYVNDGVYGSFNCILYDHAHVKPLLQKR PKPDEKYSSSIWGPTCDGLDRIVERCDLPEMHVGDWMLFENMGAYTVAASFTFNGFQ RPTIIYVMSGPAWQLMQ		
	SEQ ID NO: 123	1389 bp	
NOV10h, 258330472 DNA Sequence	ACCATGAACAACCTTTGGTAATGAAGAGTTTGACTGCCACTTCCTCGATGAAGGTTTTA CTGCCAAGGACATTCTGGACCAGAAAATTAATGAAGTTTCTTCTCTGATGATAAGGA TGCCTTCTATGTGGCAGACCTGGGAGACATTCTAAAGAAACATCTGAGGTGGTTAAAA GCTCTCCCTCGTGTCACCCCTTTTATGCAGTCAAATGTAATGATAGCAAAGCCATCG TGAAGACCTTGCTGCTACCGGACAGGATTGACTGTGCTAGCAAGACTGAAATACA GTTGGTGCAGAGTCTGGGGGTGCCTCCAGAGAGGATTATCTATGCAAATCCTTGTA CAAGTATCTCAAATTAAGTATGCTGCTAATAATGGAGTCCAGATGATGACTTTTGATA GTGAAGTTGAGTTGATGAAAGTTGCCAGAGCACATCCCAAAGCAAAGTTGGTTTTGCG GATTGCCACTGATGATTCAAAGCAGTCTGTCGTCTCAGTGTGAAATTCGGTGCCACG CTCAGAACAGCAGGCTCCTTTTGAACGGGCGAAAGAGCTAAATATCGATGTTGTTG GTGTCAGCTTCCATGTAGGAAGCGGCTGTACCGATCCTGAGACCTTCGTGACGCAAT CTCTGATGCCCGCTGTGTTTTGACATGGGGGCTGAGGTTGGTTTCAGCATGTATCTG CTTGATATTGGCGGTGGCTTCTCGGATCTGAGGATGTGAACTTAAATTTGAAGAGA TCACCGCGTAATCAACCAGCGTTGGACAAATACTTTCGTGAGCTCTGGAGTGAG AATCATAGCTGAGCCCGGCAGATACTATGTTGCATCAGCTTTCACGCTTGCAGTTAAT ATCATTGCCAAGAAAATTGTATTAAAGGAACAGACGGGCTCTGATGACGAAGATGAGT CGAGTGAGCAGACCTTATGTATTATGTGAATGATGGCGTCTATGGATCATTTAATTG CATACTCTATGACCACGCACATGTAAAGCCCTTCTGCAAAAGAGACCTAAACCAGAT GAGAAGTATTATTCATCCAGCATATGGGGACCAACATGTGATGGCCTCGATCGGATTG TTGAGCGCTGTGACCTGCCTGAAATGCATGTGGGTGATTGGATGCTCTTTGAAACAT GGGCGCTTACACTGTGTGCTGCTGCCTCTACGTTCAATGGCTTCCAGAGGCCGACGATC		

	TACTATGTGATGTCAGGGCCTGCGTGGCAACTCATGCAGCAATTCAGAACCTGACT TCCCACCCGAAGTAGAGGAACAGGATGCCAGCACCTGCTGTGTCTTGTGCCTGGGA GAGTGGGATGAAACGCCACAGAGCAGCCTGTGCTTCGGCTAGTATTAATGTGTAG		
	ORF Start: at 1		ORF Stop: TAG at 1387
	SEQ ID NO: 124	462 aa	MW at 51248.7kD
NOV10h, 258330472 Protein Sequence	TMNFGNEEFDCFLDEGFTAKDILDQKINEVSSDDKDAFYVADLGDILKKHLRWLK ALPRVTPFYAVKCNDSKAIKTLAATGTGFDCASKTEIQLVQSLGVPPERIIYANPCK QVSIKYAANNGVQMMTFDSEVELMKVARAHPKAKLVLRATDDSKAVCRLSVKFGAT LRTSRLLLERAKELNIDVVGVSFHVSGGCTDPETFVQAI SDARCVFDMGAEVGFSMYL LDIGGGFPGSEDEVKLKFEEITGVINPALDKYFSPDSGVRIIAEPGRYYVASAFTLAVN IIAKKIVLKEQTGSDDDESESEQTFMYVNDGVYGSFNCLYDHAHVKPLLQKRPKPD EKYYSSSIWGPTCDGLDRIVERCDLPEMHVGDWMLFENMGAYTVAAASTFNGFQRP TYVMGPAWQLMQQFQNPDPFPEVEEQDASTLPVSCAWESGMKRHRACASASINV		
	SEQ ID NO: 125		1386 bp
NOV10i, 258330611 DNA Sequence	CATGAACAACCTTTGGTAATGAAGAGTTTGACTGCCACTTCCTCGATGAAGGTTTTACT GCCAAGGACATTCTGGACCAGAAAATTAATGAAGTTTCTTCTCTGATGATAAGGATG CCTTCTATGTGGCAGACCTGGGAGACATTCTAAAGAAACATCTGAGGTGGTTAAAGC TCTCCCTCGTGTACCCCTTTTATGCAGTCAAATGTAATGATAGCAAAGCCATCGTG AAGACCTTGCTGTACCGGGACAGGATTTGACTGTGCTAGCAAGACTGAAATACAGT TGGTGCAGAGTCTGGGGTGCCTCCAGAGAGGATTATCTATGCAAATCCTTGTAACA AGTATCTCAAATTAAGTATGCTGCTAATAATGGAGTCCAGATGATGACTTTTGATAGT GAAGTTGAGTTGATGAAAGTTGCCAGAGCACATCCCAAAGCAAAGTTGGTTTTGCGGA TTGCCACTGATGATTCCAAAGCAGTCTGTCTGTCTCAGTGTGAAATTCGGTGCCACGCT CAGAACCAGCAGGCTCCTTTTGGAACGGGCGAAAGAGCTAAATATCGATGTTGTTGGT GTCAGCTTCCATGTAGGAACGGCTGTACCGATCCTGAGACCTTCGTGCAGGCAATCT CTGATGCCCGCTGTGTTTTGACATGGGGGCTGAGGTTGGTTTCAGCATGTATCTGCT TGATATTGGCGGTGGCTTTCTGGATCTGAGGATGTGAAACTTAAATTTGAAGAGATC ACCGGCGTAATCAACCAGCGTTGGACAAATACTTTCCGTGAGACTCTGGAGTGAGAA TCATAGCTGAGCCCGGCAGATACTATGTTGCATCAGCTTTCACGCTTGCAAGTAAATAT CATTGCCAAGAAAATTGTATTAAAGGAACAGACGGGCTCTGATGACGAAGATGAGTCG AGTGAGCAGACCTTTATGTATTATGTGAATGATGGCGTCTATGGATCATTTAATTGCA TACTCTATGACCACGCACATGTAAAGCCCTTCTGCAAAGAGACCTAAACCAGATGA GAAGTATTATTTCATCCAGCATATGGGGACCAACATGTGATGGCCTCGATCGGATTGTT GAGCGCTGTGACCTGCCTGAAATGCATGTGGGTGATTGGATGCTCTTTGAAAACATGG GCGCTTACACTGTTGCTGCTGCCTCTACGTTCAATGGCTTCCAGAGGCCGACGATCTA CTATGTGATGTCAGGGCCTGCGTGGCAACTCATGCAGCAATTCAGAACCTGACTTC CCACCCGAAGTAGAGGAACAGGATGCCAGCACCTGCCTGTGTCTTGTGCCTGGGAGA GTGGGATGAAACGCCACAGAGCAGCCTGTGCTTCGGCTAGTATTAATGTGTA		
	ORF Start: ATG at 2		ORF Stop: end of sequence
	SEQ ID NO: 126	462 aa	MW at 51147.6kD
NOV10i, 258330611 Protein Sequence	MNFNGNEEFDCFLDEGFTAKDILDQKINEVSSDDKDAFYVADLGDILKKHLRWLKA LPRVTPFYAVKCNDSKAIKTLAATGTGFDCASKTEIQLVQSLGVPPERIIYANPCKQ VSQIKYAANNGVQMMTFDSEVELMKVARAHPKAKLVLRATDDSKAVCRLSVKFGATL RTSRLLLERAKELNIDVVGVSFHVSGGCTDPETFVQAI SDARCVFDMGAEVGFSMYL DIGGGFPGSEDEVKLKFEEITGVINPALDKYFSPDSGVRIIAEPGRYYVASAFTLAVNI IAKKIVLKEQTGSDDDESESEQTFMYVNDGVYGSFNCLYDHAHVKPLLQKRPKPD KYYSSSIWGPTCDGLDRIVERCDLPEMHVGDWMLFENMGAYTVAAASTFNGFQRP TYVMGPAWQLMQQFQNPDPFPEVEEQDASTLPVSCAWESGMKRHRACASASINVX		
	SEQ ID NO: 127	1305 bp	
NOV10j, 260481330 DNA Sequence	CACCATCACCACCATCACAACAACCTTTGGTAATGAAGAGTTTGACTGCCACTTCCTCG ATGAAGGTTTTACTGCCAAGGACATTCTGGACCAGAAAATTAATGAAGTTTCTTCTTC TGATGATAAGGATGCCTTCTATGTGGCAGACCTGGGAGACATTCTAAAGAAACATCTG AGGTGGTTAAAGCTCTCCCTCGTGTACCCCTTTTATGCAGTCAAATGTAATGATA GCAAAGCCATCGTGAAGACCTTGCTGTCTACCGGCACAGGATTTGACTGTGCTAGCAA GACTGAAATACAGTTGGTGCAGAGTCTGGGGTGCCTCCAGAGAGGATTATCTATGCA		

	AATCCTTGTAACAAGTATCTCAAATTAAGTATGCTGCTAATAATGGAGTCCAGATGA TGA CTTTGTAGTAGTGAAGTTGAGTTGATGAAAGTTGCCAGAGCACATCCCAAAGCAAA GTTGGTTTTGCGGATTGCCACTGATGATTCCAAAGCAGTCTGTCTCTCAGTGTGAAA TTCGGTGCCACGCTCAGAACCAGCAGGCTCCTTTTGGAACGGGCGAAAGAGCTAAATA TCGATGTTGTTGGTGTGAGTTCCATGTAGGAAGCGGCTGTACCGATCCTGAGACCTT CGTGCAGGCAATCTCTGATGCCCGCTGTGTTTTTGACATGGGGGCTGAGGTTGTTTT AGCATGTATCTGCTTGATATTGGCGGTGGCTTTCTGGATCTGAGGATGTGAACTTA AATTGAAGAGATCACCAGCGTAATCAACCCAGCGTTGGACAAATACTTTCCGTCAGA CTCTGGAGTGAGAATCATAGCTGAGCCCGGCAGATACTATGTTGCATCAGCTTTCACG CTTGCACTTAATATCATTGCCAAGAAAATTGTATTAAAGGAACAGACGGGCTCTGATG ACGAAGATGAGTCGAGTGAGCAGACCTTTATGTATTATGTGAATGATGGCGTCTATGG ATCATTTAATTGCATACTCTATGACCACGCACATGTAAAGCCCCCTTCTGCAAAAGAGA CCTAAACCAGATGAGAAGTATTATTCATCCAGCATATGGGGACCAACATGTGATGGCC TCGATCGGATTGTTGAGCGCTGTGACCTGCCTGAAATGCATGTGGGTGATTGGATGCT CTTTGAAAACATGGGCGCTTACACTGTTGCTGCTGCCTCTACGTTCAATGGCTTCCAG AGGCCGACGATCTACTATGTGATGTCAGGGCCTGCGTGGCAACTCATGCAGTAGCGCG CCGCACTCGAGCACCACCACCACCAC		
	ORF Start: at 1		ORF Stop: TAG at 1270
	SEQ ID NO: 128	423 aa	MW at 47152.2kD
NOV10j, 260481330 Protein Sequence	HHHHHHNNFNGNEEFDCHFLDEGFTAKDILDQKINEVSSDDKDAFYVADLGDILKKHL RWL KALPRVTPFYAVKCNDSKAI VKTLAATGTGFD CASKTEIQLVQSLGVPPERIIYA NPCKQVSIKYAANNVQMMTFDSEVELMKVARAHPKAKLVLRITDDSKAVCRLSVK FGATLRTSRLLLLERAKELNIDVVGVSFHVSGSCTDPETFVQAI S DARCVFDMGAEVGF SMYLLDIGGGFPGSEDEVKLFEEITGVINPALDKYFSPDSGVRIIAEPGRYYVASAFT LAVNI IAKKIVLKEQTGSDDDEDESSEQTFMYVNDGVYGSFNCILYDHAHVKPLLQKR PKPDEKYYSSSIWGPTCDGLDRIVERCDLPEMHVGDWMLFENMGAYTVAAASTFNGFQ RPTIYYVMSPAWQLMQ		
	SEQ ID NO: 129	1416 bp	
NOV10k, CG124907-02 DNA Sequence	CGCGGATCCACCATGAACAACCTTTGGTAATGAAGAGTTGACTGCCACTTCCTCGATG AAGGTTTTACTGCCAAGGACATTCTGGACCAGAAAATTAATGAAGTTTCTTCTCTGA TGATAAGGATGCCTTCTATGTGGCAGACCTGGGAGACATTCTAAAGAAACATCTGAGG TGGTTAAAAGCTCTCCCTCGTGTACCCCTTTTATGCAGTCAAATGTAATGATAGCA AAGCCATCGTGAAGACCTTGCTGCTACCGGGACAGGATTTGACTGTGCTAGCAAGAC TGAAATACAGTTGGTGCAGAGTCTGGGGGTGCCTCCAGAGAGGATTATCTATGCAAAAT CCTTGTAACAAGTATCTCAAATTAAGTATGCTGCTAATAATGGAGTCCAGATGATGA CTTTTGATAGTGAAGTTGAGTTGATGAAAGTTGCCAGAGCACATCCCAAAGCAAAGTT GGTTTTGCGGATTGCCACTGATGATTCCAAAGCAGTCTGTCTCAGTGTGAAATTC GGTGCCACGCTCAGAACCAGCAGGCTCCTTTTGGAACGGGCGAAAGAGCTAAATATCG ATGTTGTTGGTGTGAGCTTCCATGTAGGAAGCGGCTGTACCGATCCTGAGACCTCGT GCAGGCAATCTCTGATGCCCGCTGTGTTTTTGACATGGGGGCTGAGGTTGGTTTCAGC ATGTATCTGCTTGATATTGGCGGTGGCTTTCTGGATCTGAGGATGTGAACTTAAT TTGAAGAGATCACCAGCGTAATCAACCCAGCGTTGGACAAATACTTTCCGTCAGACTC TGGAGTGAGAATCATAGCTGAGCCCGGCAGATACTATGTTGCATCAGCTTTCACGCTT GCAGTTAATATCATTGCCAAGAAAATTGTATTAAAGGAACAGACGGGCTCTGATGACG AAGATGAGTCGAGTGAGCAGACCTTTATGTATTATGTGAATGATGGCGTCTATGGATC ATTTAATTGCATACTCTATGACCACGCACATGTAAAGCCCCCTTCTGCAAAAGAGACCT AAACCAGATGAGAAGTATTATTCATCCAGCATATGGGGACCAACATGTGATGGCCTCG ATCGGATTGTTGAGCGCTGTGACCTGCCTGAAATGCATGTGGGTGATTGGATGCTCTT TGAAAACATGGGCGCTTACACTGTTGCTGCTGCCTCTACGTTCAATGGCTTCCAGAGG CCGACGATCTACTATGTGATGTGAGGGCTGCGTGGCAACTCATGCAGCAATTCAGA ACCTTGACTTCCACCCGAGTAGAGGAACAGGATGCCAGCACCTGCTGTGCTTGTG TGCTTGGGAGAGTGGGATGAAACGCCACAGAGCAGCCTGTGCTTCGGCTAGTATTAAT GTGTAGGCGGCGCTTTTTTCCTT		
	ORF Start: ATG at 13		ORF Stop: TAG at 1396
	SEQ ID NO: 130	461 aa	MW at 51147.6kD
NOV10k, CG124907-02	MNNFNGNEEFDCHFLDEGFTAKDILDQKINEVSSDDKDAFYVADLGDILKKHLRWLKA LPRVTPFYAVKCNDSKAI VKTLAATGTGFD CASKTEIQLVQSLGVPPERIIYANPCKQ		

CG124907-02 Protein Sequence	VSQIKYAANNGVQMMTFDSEVELMKVARAHPKAKLVLRITDSDSKAVCRLSVKFGATL RTSRLLLERAKELNIDVVGVSFHVSGCTDPETFVQAI SDARCVFDMGAEVGFSMYLL DIGGGFPGSESDVKLFEEITGVINPALDKYFSPDSGVRI IAEPRYYVASAFTLAVNI IAKKIVLKEQTGSDDDESESEQTFMYVNDGVYGSFNCILYDHAHVKPLLQKRPKPDE KYYSSSIWGPTCDGLDRIVERCDLPEMHVGDWMLFENMGAYTVAAASTFNGFQRPTIY YVMSGPAWQLMQQFQNPDPFPEVEEQDASTLPVSCAWESGMKRHRAACASASINV		
	SEQ ID NO: 131	1410 bp	
NOV10l, CG124907-03 DNA Sequence	ACCATGGGCCACCATCACCACCATCACAACAACCTTTGGTAATGAAGAGTTTGACTGCC ACTTCCCTCGATGAAGGTTTTACTGCCAAGGACATTCTGGACCAGAAAATTAATGAAGT TTCTTCTTCTGATGATAAGGATGCCTTCTATGTGGCAGACCTGGGAGACATTCTAAAG AAACATCTGAGGTGTTAAAGCTCTCCCTCGTGTACCCCTTTTATGCAGTCAAAT GTAATGATAGCAAAGCCATCGTGAAGACCTTGCTGCTACCGGGACAGGATTTGACTG TGCTAGCAAGACTGAAATACAGTTGGTGCAGAGTCTGGGGGTGCCTCCAGAGAGGATT ATCTATGCAAATCCTTGTAACAAGTATCTCAAATTAAGTATGCTGCTAATAATGGAG TCCAGATGATGACTTTTGATAGTGAAGTTGAGTTGATGAAAGTTGCCAGAGCACATCC CAAAGCAAAGTTGGTTTTGCGGATTGCCACTGATGATTCAAAGCAGTCTGTCTCTC AGTGTGAAATTCGGTGCCACGCTCAGAACAGCAGGCTCCTTTTGGAACGGGCGAAAG AGCTAAATATCGATGTTGTTGGTGTGAGCTTCCATGTAGGAAGCGGCTGTACCGATCC TGAGACCTTCGTGCAGGCAATCTCTGATGCCCGCTGTGTTTTGACATGGGGGCTGAG GTTGGTTTCAGCATGTATCTGCTTGATATTGGCGGTGGCTTTCCTGATCTGAGGATG TGAAACTTAAATTTGAAGAGATCACCGGCGTAATCAACCCAGCGTTGGACAAATACTT TCCGTCAGACTCTGGAGTGAGAATCATAGCTGAGCCCGGCAGATACTATGTTGCATCA GCTTTCACGCTTGCAAGTAAATATCATTGCCAAGAAAATGTATTAAAGGAACAGACGG GCTCTGATGACGAAGATGAGTCGAGTGAGCAGACCTTTATGTATTATGTGAATGATGG CGTCTATGGATCATTAAATTGCATACTCTATGACCACGCACATGTAAGCCCTTCTG CAAAGAGACCTAAACCAGATGAGAAGTATTATTATCCAGCATATGGGGACCAACAT GTGATGGCCTCGATCGGATTGTTGAGCGCTGTGACCTGCCTGAAATGCATGTGGGTGA TTGGATGCTCTTTGAAAACATGGGCGCTTACACTGTTGCTGCTGCCTTACGTTCAAT GGCTTCCAGAGGCCGACGATCTACTATGTGATGTGAGGCTGCGTGGCAACTCATGC AGCAATTCAGAACCCCTGACTTCCCACCCGAAGTAGAGGAACAGGATGCCAGCACCT GCCTGTGCTTGTGCTGGGAGAGTGGGATGAAACGCCACAGAGCAGCCTGTGCTTCG GCTAGTATTAATGTGTAG		
	ORF Start: at 1		ORF Stop: TAG at 1408
	SEQ ID NO: 132	469 aa	MW at 52128.6kD
NOV10l, CG124907-03 Protein Sequence	TMGHHHHHHNNFNGNEEFDFHFLDEGFTAKDILDQKINEVSSDDKDAFYVADLGDILK KHLRWL KALPRVTPFYAVKCNDSKAI VKTLAATGTGFDCASKTEIQLVQSLGVPPERI IYANPCKQVSQIKYAANNGVQMMTFDSEVELMKVARAHPKAKLVLRITDSDSKAVCR SVKFGATLRTSRLLLERAKELNIDVVGVSFHVSGCTDPETFVQAI SDARCVFDMGA VGFSMYLLDIGGGFPGSESDVKLFEEITGVINPALDKYFSPDSGVRI IAEPRYYVAS AFTLAVNI IAKKIVLKEQTGSDDDESESEQTFMYVNDGVYGSFNCILYDHAHVKPLL QKRPKPDEKYYSSSIWGPTCDGLDRIVERCDLPEMHVGDWMLFENMGAYTVAAASTFN GFQRPTIYYVMSGPAWQLMQQFQNPDPFPEVEEQDASTLPVSCAWESGMKRHRAACAS ASINV		
	SEQ ID NO: 133	1407 bp	
NOV10m, CG124907-04 DNA Sequence	ACCATGAACAACCTTTGGTAATGAAGAGTTTGACTGCCACTTCTCGATGAAGGTTTTA CTGCCAAGGACATTCTGGACCAGAAAATTAATGAAGTTTCTTCTTCTGATGATAAGGA TGCCTTCTATGTGGCAGACCTGGGAGACATTCTAAAGAAACATCTGAGGTGGTTAAAA GCTCTCCCTCGTGTACCCCTTTTATGCAGTCAAATGTAATGATAGCAAAGCCATCG TGAAGACCTTGCTGCTACCGGACAGGATTTGACTGTGCTAGCAAGACTGAAATACA GTTGGTGCAGAGTCTGGGGGTGCCTCCAGAGAGGATTATCTATGCAAATCCTTGTA CAAGTATCTCAAATTAAGTATGCTGCTAATAATGGAGTCCAGATGATGACTTTTGATA GTGAAGTTGAGTTGATGAAAGTTGCCAGAGCACATCCCAAAGCAAAGTTGGTTTTGCG GATTGCCACTGATGATTCAAAGCAGTCTGTCTGCTCAGTGTGAAATTCGGTGCCACG CTCAGAACAGCAGGCTCCTTTTGAACGGGCGAAAGAGCTAAATATCGATGTTGTTG GTGTGAGCTTCCATGTAGGAAGCGGCTGTACCGATCCTGAGACCTTCGTGCAGGCAAT CTCTGATGCCCCGTGTGTTTTGACATGGGGGCTGAGGTGGTTTTGAGCATGTATCTG CTTGATATTGGCGGTGGCTTTCCTGGATCTGAGGATGTGAAACTTAAATTTGAAGAGA		

	TCACCGGCGTAATCAACCCAGCGTTGGACAAATACTTTCCGTCAGACTCTGGAGTGAG AATCATAGCTGAGCCCGGCAGATACTATGTTGCATCAGCTTTACGCTTGCAGTTAAT ATCATTGCCAAGAAAATTGTATTAAAGGAACAGACGGGCTCTGATGACGAAGATGAGT CGAGTGAGCAGACCTTTATGTATTATGTGAATGATGGCGTCTATGGATCATTAAATTTG CATACTCTATGACCACGCACATGTAAAGCCCTTCTGCAAAAGAGACCTAAACCAGAT GAGAAGTATTATTATCCAGCATATGGGGACCAACATGTGATGGCCTCGATCGGATTG TTGAGCGCTGTGACCTGCCTGAAATGCATGTGGGTGATTGGATGCTCTTTGAAACAT GGGCGCTTACACTGTTGCTGCTGCCTCTACGTTCAATGGCTTCCAGAGGCCGACGATC TACTATGTGATGTCAGGGCCTGCGTGGCAACTCATGACGAATTCAGAACCCTGACT TCCCACCCGAAGTAGAGGAACAGGATGCCAGCACCTGCCTGTGTCTTGTGCCTGGGA GAGTGGGATGAAACGCCACAGAGCAGCCTGTGCTTCGGCTAGTATTAATGTGCACCAT CACCACCATCACTGA		
	ORF Start: at 1		ORF Stop: TGA at 1405
	SEQ ID NO: 134	468 aa	MW at 52071.6kD
NOV10m, CG124907-04 Protein Sequence	TMNFGNEEFDCFLDEGFTAKDILDQKINEVSSSDDKDAFYVADLGDILKKHLRWL ALPRVTPFYAVKCNDSKAIVKTLAATGTGFDCAKTEIQLVQSLGVPPERIIYANPCK QVSQIKYAANNGVQMFTFDSEVELMKVARAHPKAKLVLRATDDSKAVCRLSVKFGAT LRTSRLLLERAKELNIDVVGVSFHVSGCTDPETFVQAI SDARCVFDMGAEVGFSMYL LDIGGGFPGSEDEVKLKFEETGVINPALDKYFSDSGVRIIAEPGRYVVASAFTLAVN IIAKKIVLKEQTGSDDDESESEQTFMYVNDGVYGSFNCILYDHAHVKPLLQKRPKPD EKYYSSSIWGPTCDGLDRIVERCDLPEMHVGDWMLFENMGAYTVAAASTFNGFQRP TIYYVMSGPAWQLMQQFQNPDPFPEVEEQDASTLPVSCAWESGMKRHRAACASASINVH HHHH		
	SEQ ID NO: 135	1305 bp	
NOV10n, CG124907-05 DNA Sequence	ACATCATCACCACCATCAACAACCTTTGGTAATGAAGAGTTTGACTGCCACTTCCTCG ATGAAGGTTTTACTGCCAAGGACATTCTGGACCAGAAAATTAATGAAGTTTCTTCTTC TGATGATAAGGATGCCTTCTATGTGGCAGACCTGGGAGACATTCTAAAGAAACATCTG AGGTGGTTAAAGCTCTCCCTCGTGTACCCCTTTTATGCAGTCAAATGTAATGATA GCAAAGCCATCGTGAAGACCTTGCTGTACCGGGACAGGATTTGACTGTGCTAGCAA GACTGAAATACAGTTGGTGCAGAGTCTGGGGGTGCCTCCAGAGAGGATTATCTATGCA AATCCTTGTAACAAGTATCTCAAATTAAGTATGCTGCTAATAATGAAGTCCAGATGA TGACTTTTGATAGTGAAGTTGAGTTGATGAAAGTTGCCAGAGCACATCCCAAAGCAAA GTTGGTTTTGCGGATTGCCACTGATGATTCAAAGCAGTCTGTCTCTCAGTGTGAAA TTCCGTGCCACGCTCAGAACCAGCAGGCTCCTTTTGAACGGGGCAAAGAGCTAAATA TCGATGTTGTTGGTGTGAGCTTCCATGTAGGAAGCGGCTGTACCGATCCTGAGACCTT CGTGCAGCAATCTCTGATGCCCGCTGTGTTTTGACATGGGGGTGAGGTTGGTTTC AGCATGTATCTGCTTGATATTGGCGGTGGCTTTCCCTGGATCTGAGGATGTGAACCTTA AATTGAAGAGATCACCGCGTAATCAACCCAGCGTTGGACAAATACTTTCCGTCAGA CTCTGGAGTGAGAATCATAGCTGAGCCCGGCAGATACTATGTTGCATCAGCTTTCAG CTTGCAGTTAATATCATTGCCAAGAAAATTGTATTAAAGGAACAGACGGGCTCTGATG ACGAAGATGAGTCGAGTGAGCAGACCTTTATGTATTATGTGAATGATGGCGTCTATGG ATCATTTAATTGCATACTCTATGACCACGCACATGTAAAGCCCTTCTGCAAAAGAGA CCTAAACCAGATGAGAAGTATTATTCATCCAGCATATGGGGACCAACATGTGATGGCC TCGATCGGATTGTTGAGCGCTGTGACCTGCCTGAAATGCATGTGGGTGATTGGATGCT CTTTGAAAACATGGGCGCTTACACTGTGCTGCTGCTCTACGTTCAATGGCTTCCAG AGGCCGACGATCTACTATGTGATGTGAGGCGCTGCGTGGCAACTCATGCAGTAGGCGG CCGCACTCGAGCACCACCACCACCAC		
	ORF Start: at 1		ORF Stop: TAG at 1270
	SEQ ID NO: 136	423 aa	MW at 46885.9kD
NOV10n, CG124907-05 Protein Sequence	TSSPPSNNFGNEEFDCFLDEGFTAKDILDQKINEVSSSDDKDAFYVADLGDILKKHL RWLKAALPRVTPFYAVKCNDSKAIVKTLAATGTGFDCAKTEIQLVQSLGVPPERIIYA NPCKQVSQIKYAANNGVQMFTFDSEVELMKVARAHPKAKLVLRATDDSKAVCRLSVK FGATLRTSRLLLERAKELNIDVVGVSFHVSGCTDPETFVQAI SDARCVFDMGAEVGF SMYLLDIGGGFPGSEDEVKLKFEETGVINPALDKYFSDSGVRIIAEPGRYVVASAFT LAVNIIAKKIVLKEQTGSDDDESESEQTFMYVNDGVYGSFNCILYDHAHVKPLLQKR PKPDEKYYSSSIWGPTCDGLDRIVERCDLPEMHVGDWMLFENMGAYTVAAASTFNGFQ RPTIYYVMSGPAWQLMQ		

	SEQ ID NO: 137	1305 bp	
NOV10o, CG124907-06 DNA Sequence	CACCATCACCACCATCACAACTTTGGTAATGAAGAGTTTGAAGTCCACTTCCTCG ATGAAGGTTTTACTGCCAAGGACATTCTGGACCAGAAAATTAATGAAGTTTCTTCTTC TGATGATAAGGATGCCTTCTATGTGGCAGACCTGGGAGACATTCTAAAGAAACATCTG AGGTGGTTAAAGCTCTCCCTCGTGTACCCCTTTTATGCAGTCAAATGTAATGATA GCAAAGCCATCGTGAAGACCTTGCTGCTACCGGGACAGGATTTGACTGTGCTAGCAA GACTGAAATACAGTTGGTGCAGAGTCTGGGGTGCCTCCAGAGAGGATTATCTATGCA AATCCTTGTAACAAGTATCTCAAATTAAGTATGCTGCTAATAATGGAGTCCAGATGA TGACTTTTGATAGTGAAGTTGAGTTGATGAAAGTTGCCAGAGCACATCCCAAAGCAAA GTTGGTTTTCGGATTGCCACTGATGATTCAAAGCAGTCTGTCTGCTCAGTGTGAAA TTCGGTGCCACGCTCAGAACCAGCAGGCTCCTTTTGGAAACGGGCGAAAGAGCTAAATA TCGATGTTGTTGGTGTGAGCTTCCATGTAGGAAGCGGCTGTACCGATCCTGAGACCTT CGTGCAGGCAATCTCTGATGCCGCTGTGTTTGTGACATGGGGGCTGAGGTTGGTTTC AGCATGTATCTGCTTGATATTGGCGGTGGCTTTCCTGGATCTGAGGATGTGAACTTA AATTTGAAGAGATCACCGGCGTAATCAACCAGCGTTGGACAAATACTTTCCGTCAGA CTCTGGAGTGAGAATCATAGCTGAGCCCGGCAGATACTATGTTGCATCAGCTTTCACG CTTGCACTTAATATCATTGCCAAGAAAATGTATTAAAGGAACAGACGGGCTCTGATG ACGAAGATGAGTCGAGTGAGCAGACCTTTATGTATTATGTAATGATGGCCTCTATGG ATCATTTAATTGCATACTCTATGACCACGCACATGTAAAGCCCTTCTGCAAAAGAGA CCTAAACCAGATGAGAAGTATTATTATCCAGCATATGGGGACCAACATGTGATGGCC TCGATCGGATTGTTGAGCGCTGTGACCTGCCTGAAATGCATGTGGGTGATTGGATGCT CTTTGAAAACATGGGCGCTTACACTGTTGCTGCTGCCTCTACGTTCAATGGCTTCCAG AGGCCGACGATCTACTATGTGATGTGAGGCTGCGTGGCAACTCATGCAGTAGGCGG CCGCACTCGAGCACCAACCACCACCAC		
	ORF Start: at 19		ORF Stop: TAG at 1270
	SEQ ID NO: 138	417 aa	MW at 46329.3kD
NOV10o, CG124907-06 Protein Sequence	NNFGNEEFDCFLDEGFTAKDILDQKINEVSSSDDKDAFYVADLGDILKKHLRWLKAL PRVTPFYAVKCNDSKAIIVKTLAATGTGFDCASKTEIQLVQSLGVPPERIIYANPCKQV SQIKYAANNGVQMMTFDSEVELMKVARAHPKAKLVLRITDDSKAVCRLSVKFGATLR TSRLLLERAKELNIDVVGVSFHVSGCTDPETFVQAI SDARCVFDMGAEVGFMSYLLD IGGGFPGSEDVKLKFEETITGVINPALDKYFPSDSGVR IIAEPGRYYVASAFTLAVNII AKKIVLKEQTGSDEDESSEQTFMYVNDGVYGSFNCILYDHAHVKPLLQKRPKPDEK YYSSSIWGPTCDGLDRIVERCDLPEMHVGDWMLFENMGAYTVAAASTFNGFQRPRTIYY VMSGPAWQLMQ		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 10B.

Table 10B. Comparison of NOV10a against NOV10b through NOV10o.		
Protein Sequence	NOV10a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV10b	1..461 1..461	461/461 (100%) 461/461 (100%)
NOV10c	1..461 5..465	461/461 (100%) 461/461 (100%)
NOV10d	2..461 10..469	460/460 (100%) 460/460 (100%)
NOV10e	1..461 2..462	461/461 (100%) 461/461 (100%)
NOV10f	2..461	460/460 (100%)

	7..466	460/460 (100%)
NOV10g	2..418 7..423	417/417 (100%) 417/417 (100%)
NOV10h	1..461 2..462	461/461 (100%) 461/461 (100%)
NOV10i	1..461 1..461	461/461 (100%) 461/461 (100%)
NOV10j	2..418 7..423	417/417 (100%) 417/417 (100%)
NOV10k	1..461 1..461	461/461 (100%) 461/461 (100%)
NOV10l	2..461 10..469	460/460 (100%) 460/460 (100%)
NOV10m	1..461 2..462	461/461 (100%) 461/461 (100%)
NOV10n	2..418 7..423	417/417 (100%) 417/417 (100%)
NOV10o	2..418 1..417	417/417 (100%) 417/417 (100%)

Further analysis of the NOV10a protein yielded the following properties shown in Table 10C.

Table 10C. Protein Sequence Properties NOV10a	
PSort analysis:	0.6000 probability located in nucleus; 0.3922 probability located in microbody (peroxisome); 0.1000 probability located in mitochondrial matrix space; 0.1000 probability located in lysosome (lumen)
SignalP analysis:	No Known Signal Sequence Predicted

5 A search of the NOV10a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 10D.

Table 10D. Geneseq Results for NOV10a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV10a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAG73867	Human colon cancer antigen protein SEQ ID NO:4631 - Homo	1..461 6..466	461/461 (100%) 461/461 (100%)	0.0

	sapiens, 466 aa. [WO200122920-A2, 05-APR-2001]			
AAB58391	Lung cancer associated polypeptide sequence SEQ ID 729 - Homo sapiens, 466 aa. [WO200055180-A2, 21-SEP-2000]	1..461 6..466	461/461 (100%) 461/461 (100%)	0.0
AAR37270	ODC - Synthetic, 461 aa. [EP542287-A, 19-MAY-1993]	1..461 1..461	460/461 (99%) 461/461 (99%)	0.0
AAB52181	Human secreted protein BLAST search protein SEQ ID NO: 137 - Homo sapiens, 428 aa. [WO200061624-A1, 19-OCT-2000]	17..444 1..428	427/428 (99%) 428/428 (99%)	0.0
AAW76000	Ornithine decarboxylase amino acid sequence - Mus sp, 461 aa. [US5811634-A, 22-SEP-1998]	1..461 1..461	417/461 (90%) 434/461 (93%)	0.0

In a BLAST search of public sequence databases, the NOV10a protein was found to have homology to the proteins shown in the BLASTP data in Table 10E.

Table 10E. Public BLASTP Results for NOV10a				
Protein Accession Number	Protein/Organism/Length	NOV10a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
P11926	Ornithine decarboxylase (EC 4.1.1.17) (ODC) - Homo sapiens (Human), 461 aa.	1..461 1..461	461/461 (100%) 461/461 (100%)	0.0
P27117	Ornithine decarboxylase (EC 4.1.1.17) (ODC) - Bos taurus (Bovine), 461 aa.	1..461 1..461	431/461 (93%) 444/461 (95%)	0.0
P09057	Ornithine decarboxylase (EC 4.1.1.17) (ODC) - Rattus norvegicus (Rat), 461 aa.	1..461 1..461	422/461 (91%) 434/461 (93%)	0.0
P27119	Ornithine decarboxylase (EC 4.1.1.17) (ODC) - Mus pahari (Shrew mouse), 461 aa.	1..461 1..461	421/461 (91%) 436/461 (94%)	0.0
P00860	Ornithine decarboxylase (EC 4.1.1.17) (ODC) - Mus musculus (Mouse), 461 aa.	1..461 1..461	417/461 (90%) 434/461 (93%)	0.0

PFam analysis predicts that the NOV10a protein contains the domains shown in the Table 10F.

Table 10F. Domain Analysis of NOV10a			
Pfam Domain	NOV10a Match Region	Identities/ Similarities for the Matched Region	Expect Value
Om_Arg_deC_N	44..282	131/289 (45%) 225/289 (78%)	7.8e-132
Om_DAP_Arg_deC	285..409	68/199 (34%) 119/199 (60%)	5.6e-62

**Example 11.**

The NOV11 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 11A.

Table 11A. NOV11 Sequence Analysis			
	SEQ ID NO: 139	994 bp	
NOV11a, CG128347-01 DNA Sequence	CACACAAGTCCGCCTATGTACTCTCTGGATCGAATATTTGCTGGATTTGGAACACGAA GTCAGATGCTGTTGGGTACATAGAAGAACAGATAAGGTCCTCCACTGCCAATTTTC TGATAACAGTGATGATGAAGAATCAGAAGGCCAGACTCTGTTTGTCCCTTGTGAATTGAGTGATA AGTCGTTTCATGGATTTCAGAAGCCAGACTCTGTTTGTCCCTTGTGAATTGAGTGATA CTCAGGATGAAACACAAAAGTCAGATTTGGAGAATGAAGATTTAAAGATTGATTGTCT CCAGGAGAGTCAAGAATTGAATTTGCAAAAATTAAAGAATTCAGAACGCATACTTACT GAAGCTAAACAAAAATGAGAGAACCTACAATTAACATCAAGATGAAGGAAGATCTGA TTAAAGAATTAATAAAAAACAGGTAATGATGCCAAGTCTGTAAGCAAGCAGTATTCTTT GAAAGTAACAAAGCTAGAGCATGATGCAGAACAGGCAGCAAAAGTCAACTAAGTGAACAA CAAAAGCAGCTACAGGAGCTGGAACAAAGATCTTTCTGATGTTGCAATGAAGGTAA AATTACAGAAAGAGTTTCGTAAGAAAGATGGATGCTGCAAGCTGAGAGTTCAGGTCTT ACAGAAGAAGCAACAAGATAGTAAGAACTGGCATCACTGTCAATCCAAAATGAGAAA CGTGCTAATGAACTAGAGCAGAGTGTAGATCACATGAAATATCAAAAGATACAGCTAC AAAGAAAACCTACAAGAAGAAAATGAAAAAGGAAGCAACTGGATGCAGTAATTAAGCG GGACCAGCAAAAATCAAAGTAATATTGTCTATACATTCTGTCTAAGTATAATATGAAA TGTTAAACGGCTCAGAGCTAACGAATCCATGGTCTTCATTCAAGTTGGCTTGTGAAGTA TCTATCCTTGACTTGCCCTTCACTGCTGTCCTTATTCATTAAAGCTTTGTTCATCT ACATAGTA		
	ORF Start: ATG at 16		ORF Stop: TAA at 874
	SEQ ID NO: 140	286 aa	MW at 33507.0kD
NOV11a, CG128347-01 Protein Sequence	MYSLDRI FAGFRTRSQMLLGHIEEQDKVLHCQFSDNSDDEESEGQEKSGTRCSRSWI QKPDVSVCSLVELSDTQDETQKSDLENDLKIDCLQESQELNLQKLKNSERILTEAKQK MRELTINIKMKEDLIKELIKTGNDAKSVSKQYSLKVTKLEHDAEQAKVELTETQKQLQ ELENKDLSDVAMKVKLQKEFRKKMDAÁKLRVQVLQKKQQDSKKLASLSIQNEKRANEL EQSVDHMKYQKIQLQRKLQENEKRKQLDAVIKRDOQKIKVILSYIPAKYNMCK		
	SEQ ID NO: 141	4622 bp	
NOV11b, CG128347-02 DNA Sequence	AGGAGTCCAGCGCTCGCCGACAGGGGCTGGGCTGTCCCAGCCGGAATCCAGATCTT ACATAAGATGGAAGTCTCTCACACTAGATACTGAACATTAAATAGAAAATCTATTTAG TAAATCTAAGTTGCCATGGAAGAAAATACCAGTAAAAGTTGCTGTAAGAATTAGACCT CTGCTTTGCAAGAAGCTCTTCAATATCATCAAGTTTGTGTGAGAGTTATTCCAAACA GCCAGCAAGTTATCATTGGGAGAGATAGAGTCTTCACTTTTGATTTTGTGGCAAA AAATTCCACTCAAGATGAAGTTTATAACACATGTATAAAGCCCCTAGTGTGTCACTC ATTGAGGGCTATAATGCAACTGTTTTTGCCTATGGACAACTGGATCTGGGAAGACAT ACACCATGGAGGGGGCCATATTGCTTCAGTTGTGGAGGGCCAAAAGGGTATCATTC TCGAGCTATTCAAGAAATATTTCAAAGCATCTCTGAACATCCTAGCATTGACTTTAAT		

GTAAAAGTATCTTATATAGAAGTGTACAAGGAAGACCTAAGAGATCTTCTAGAATTGG  
 AGACATCCATGAAGGATCTTCACATCCGAGAAGATGAAAAAGGAAACACAGTGATTGT  
 TGGGGCCAGGAATGCCATGTGGAGAGTGCAGGTGAAGTGATGAGTCTTTTGGAGATG  
 GGAATGCAGCCAGACATACAGGTACCACTCAAATGAATGAGCAGCTCCAGCAGATCAC  
 ATGCAATTTTACAATCAGCATTTTGTCAAGTTTATAAAAAATATGGAGGCAGCTGAAGA  
 TGGATCATGGTATTCCCTCGGCATATTGTCTCAAAGTTCCACTTTGTGGATTGGCA  
 GGATCAGAAAGAGTAACCAAAACGGGAATACTGGTGAACGGTTCAAAGAATCCATTTC  
 AAATCAATAGTGGATTGCTGGCTTTAGGAAATGTAATAAGCGCTCTTGGGGACCCACG  
 CAGGAAGAGTTTACATATTCCATATAGGGATGCTAAAATTACCCGGCTTCTGAAAGAT  
 TCTCTGGGAGGCAGTGTAAAGACTGTCATGATCAGATGTGTGTCAGCCCTCTCTCGA  
 ATTTTGATGAGTCCCTTAAATTCTCTCAAATATGCCAACAGAGCACGGAACATTAGAAA  
 CAAACCCACTGTAACTTCAGCCCCGAGTCAGACCGTATAGATGAAATGGAATTTGAG  
 ATTAAATTGCTTCGAGAAGCTTTGCAAAGCCAGCAGGCTGGTGTGAGCCAAACTACCC  
 AGATCAATCGAGAAGGGAGTCCGTGATACAAATAGGATTTCATTCTCTTGAGGAGCAAGT  
 AGCTCAGCTTCAAGGAGAATGTCTGGGTACCAGTGTGTGTAGAAGAAGCCTTTACC  
 TTCTTGGTTGACCTAAAAGATACTGTGAGACTAAACGAAAAGCAGCAACACAACTGC  
 AGGAGTGGTTTAAACATGATCCAAGAGGTGAGGAAGGCTGTCTCACCTCATTTTCGAGG  
 AATCGGAGGCACTGCAAGTCTGGAAGAAGGACCACAGCATGTTACAGTTCTCCAGCTG  
 AAGAGAGAGCTTAAGAAATGCCAGTGTGTGCTTGTCTGATGAAGTAGTATTTAATC  
 AGAAGGAAGTGGAGGTGAAGGAATGAAGAATCAAGTGCAGATGATGGTACAGGAAAA  
 CAAAGGGCATGCTGTATCTTTGAAAGAAGCGCAAAAAGTGAATAGACTGAGAATGAA  
 AAAATAATAGAACAACAACCTTCTTGTGGATCAACTGAGTGAAGAATAACAAAACCTTA  
 ACCTGTCAGTGACTTCTTCAGCTAAAGAAAATTGTGGAGATGGGCCAGATGCCAGGAT  
 CCTTGAAGGAGACCATATACTGTACCATTTGATACTCATTTGGGGCATATATTTAT  
 ATCCCATCAAGACAAGATTCCAGGAAGGTCCACACAAGTCCGCCTATGTACTCTCTGG  
 ATCGAATATTTGCTGGATTTGCAACACGAAGTCAGATGCTGTTGGGTACATAGAAGA  
 ACAAGATAAGGTCCCTCCACTGCCAATTTTCTGATAACAGTGTATGAAGAATCAGAA  
 GGCCAAGAGAAATCTGGAAGTATAGTGAAGTCTGTTATGGATTTCAGAAGCCAGACT  
 CTGTTTGTTCCTTGTGTAATTGAGTGATACTCAGGATGAAACACAAAAGTCAGATTT  
 GGAGAATGAAGATTTAAAGATTGATTGTCTCCAGGAGAGTCAAGAATTGAATTTGCAA  
 AAATTAAAGAAATTCAGAACGCATACTTACTGAAGCTAAACAAAAATGAGAGAATTA  
 CAATTAACATCAAGATGAAGGAAGATCTGATTAAAGAATTAATAAAAACAGGAATGA  
 TGCCAAGTCTGTAAGCAAGCAGTATTCTTTGAAAGTAACAAAGCTAGAGCATGATGCA  
 GAACAGGCAAAAGTCGAAGTGAATGAAACACAAAAGCAGCTACAGGAGCTGGAAAACA  
 AAGATCTTTCTGATGTTGCAATGAAGGTAAAATTACAGAAAGAGTTTCGTAAAAAGAT  
 GGATGCTGCAAGCTGAGAGTTTCAAGTCTTGCAGAAGAAGCAACAAGATAGTAAGAAA  
 CTGGCATCACTGTCAATCCAAATGAGAAACGTGCTAATGAGCTAGAGCAGAGTGTAG  
 ATCAGATGAAATATCAAAAGATACAGCTACAAAGAAAACCTACGAGAAGAAAATGAAAA  
 AAGGAAGCAACTGGATGCAGTAATTAAGCGGGACCAGCAAAAATCAAAGTAATACAA  
 TTA AAAACAGGACAGGAAGAAGGTCTAAACCGAAAGCTGAGGACCTTGATGCATGTA  
 ACTTGAAGAGGAGAAAAGGTTCTGTTTGAAGTATAGACCATCTCCAGAAATGGATGA  
 GCAAAAGAAATGGTTAGATGAAGAAGTAGAGAAAGTTCTGAACCAACGCCAAAGTTA  
 GAGGAGCTGGAAGCAGACTTAAAGAAACGGGAGGCCATAGTTTCTAAGAAGGAGGCTC  
 TGTTACAGGAGAAGAGTCACCTGGAAAATAAGAAATTGAGATCTAGTCAGGCCCTTAAA  
 CACAGATAGTTTGAAAATATCAACTCGCCTGAACCTTACTGGAACAAGAGTTGTCTGAA  
 AAGAATGTGCAGCTCCAGACCAGTACAGCTGAGGAGAAAACAAAGATTTCAGAACAAAG  
 TTGAAGTCTCTCAGAAAAGAAAAGGATCAGCTCCAGAAACGAGACAGATGTGGATGA  
 AAAACTTAAAAATGGTAGAGTGTATCACCTGAAGAAGAATCATGTTCTTTTCCAACCTT  
 GAAGAAGGGATAGAAGCTTTGGAAGCTGCAATTGAATACAGGAATGAAAGTATCCAGA  
 ATCGCCAGAAGTCACTTAGAGCATCATTCCATAACCTCTCTCGTGGTGAAGCAAATGT  
 CTTGGAAAAGCTAGCTTGCCCTGAGTCTGTTGAGATTAGAATATTCTTTTCAGATAT  
 TTCAATAAGGTGGTGAATTTGCGAGAAGCTGAACGGAACAAACAGTTATATAATGAAG  
 AAATGAAAATGAAAGTTCTGGAACGGGATAATATGGTTCTGTAAGTATGAACTGCACCT  
 GGACCATCTAAAATTGCAGTGTGACCGGAGACTGACCCTCCAGCAAAAGGAAACACGAA  
 CAAAAGATGCAGTTGCTATTACATCATTTCAAAGAACAAGATGGAGAAGGCATTATGG  
 AAACCTTCAAACATATGAAGATAAAATCCAGCAGTTGGAAAAGATCTTTATTTCTA  
 TAAGAAAACAGCCGGGATCATAGAAGAACTTAAGGAAGTGGTAGGGGAAGCAATT  
 CGGCGGCAACTAGCATCATCAGAGTATCAAGAGGCTGGAGATGGAGTCTTGAAGCCAG  
 AAGGAGGAGGCATGCTTTCAGAAGAATTAATAATGGGCATCCAGACCTGAAAGTATGAA  
 ATTAAGTGAAGAGAAAAGAAAATGGACAGTTTCAGCAAGCAGCTTAAGAACACAGCCA  
 AATCCTCAAAGCTCTGGGAAGATATCCAGAATTACCTCCAATTCATAGTTCTTTAG

	CACCCCCCAGTGGGCATATGTTAGGTAATGAGAATAAAACAGAAACAGATGATAATCA GTTTACAAAATCTCACAGTCGACTGTCATCCCAAATTCAGGTTGTGGGAAATGTGGGA CGACTTCATGGTGTACACCTGTAAACTGTGTGCGAAAAGAATTACGTCAAATTTCCG CCTTGGAACTATCATTGCGACGTTCCAGTCTTGGAGTTGGCATTGGATCAATGGCTGC TGATTCCATCGAAGTATCTAGGAAACCAAGGGACTTAAAAACTTAGACATTGAATAAT AGAACTTTTAGTAGATATGTAAGAAATTCCTTTTCTAACCTGTTAAAACTAAAGC TCAAGTTCACCTACCTCTTTCTCTCAGAATAAAGGAAGAAGGGAGGAAGGAATCCCTAA TTCTTTTATATGCTATAGATGTGTACATCTTCTATATATATTGGGGAGTTTTAGTTT ATATTCCTCATAGTAATCAAACATGTTTTCCAATACTTGATAACATTTAAATATTATA AATACGCTTAAATGTTTTTCCAGGCATATTTGAAGATTAA		
	ORF Start: ATG at 133		ORF Stop: TAG at 4336
	SEQ ID NO: 142	1401 aa	MW at 160242.6kD
NOV11b, CG128347-02 Protein Sequence	MEEIPVKVAVRIRPLLCKEALHNNQVCVRVIPNSQQVIIGRDRVFTFDVFVGKNSTQD EVYNTCIKPLVLSLIEGYNATVFAYGQTGSGKTYTIGGGHIASVVEGQKGIIPRAIQE IFQSISEHPSIDFNVKVSIEVYKEDLRDLLELETSMKDLHIREDEKGNVTIVGAKEC HVESAGEVMSLLEMGNARHTGTTQMNEHSSSRSHAFITISICQVHKMNEAABDGWSY PRHIVSKFHFVDLAGSERVTKTGNTGERFKESIQINSGLLAGNVISALGDPRRKSSH IPYRDAKITRLKDSLGSAAKTVMITCVSPSSSNFDESLNSLYANRARNIRNKPTVN FSPESDRIDEMEFKILREALQSQQAGVSQTTQINREGSPDNRHISLEEQVAQLQG ECLGYQCCVEEAFTFLVDLKDTRVRLNEKQQHKLQEWFNMIQEVRAVLTSFRGIGGTA SLEEGPQHVTVLQLKRELKKQCQVLADEVVFNQKELEVKELKNQVQMMVQENKHAV SLKEAQKVNRLQNEKIEQQLLDQLSEELTKLNLVTSACKENCGDGPDAIPERRP YTVPFDTLGHYIYIPSRQDSRKVHTSPPMYSLDRIFAGFRTRSQMLLGHIEEQDKVL HCQFSDNSDDESEGEKSGTRCSRSWIQKPDVCSLVELSDTQDETQKSDLENEDL KIDCLQESQELNLQKLKNSERILTEAKQKMRELITINIKMEDLIKELIKTGNDKSVS KQYSLKVTLEHDAEQAKVELIETQKQLQELNKLSDVAMKVKLQKEFRKKMDAAKL RVQVLQKKQDQSKKLASLSIQNEKRANELEQSVDHMKYQKIQLQRKLREENEKRKOLD AVIKRDQKIKVIQLKTGQEEGLKPKAEDLDACNLKRRKGSFGSIDHLQKLDEQKKWL DEEVEKVLNQRQEELEADLKKREAIVSKKEALLQEKSHLENKKLRSSQALNTDSLK ISTRNLNLEQELSEKNVQLQTSTAEETKISEQVEVLQKEKDQLQRRRHVDDEKLKNG RVLSPEEEHVLFLQEEGIEALEAAIEYRNESIQRQKSLRASFNLSRGEANVLEKLA CLSPVEIRITILFRYFNKVNLREAERKQQLYNEEMKMKVLERDNMVRLESALDHLKL QCDRRLTLQKEHEQKMQLLHFFKEQDGEIGIMETFKTYEDKIQLEKDLYFYKKTSR DHKKKLKELVGEAIRQLASSEYQEAGDGVLPKPEGGMLEELKWSRPESMKLSGRE REMDSSASSLRTQPNPQKLWEDIPELPIHSSSLAPSGHMLGNENKTETDDNQFTKSH SRLSSQIQVGVNVRGLHGVTPVKLCRKELRQISALELSLRRSSLGVGIGSMAADSIEV SRKPRDLKT		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 11B.

Table 11B. Comparison of NOV11a against NOV11b.		
Protein Sequence	NOV11a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV11b	1..274 610..883	272/274 (99%) 273/274 (99%)

Further analysis of the NOV11a protein yielded the following properties shown in Table 11C.

Table 11C. Protein Sequence Properties NOV11a	
PSort	0.5517 probability located in mitochondrial matrix space: 0.3000 probability

analysis:	located in microbody (peroxisome); 0.2717 probability located in mitochondrial inner membrane; 0.2717 probability located in mitochondrial intermembrane space
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV11a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 11D.

Table 11D. Geneseq Results for NOV11a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV11a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAB42353	Human ORFX ORF2117 polypeptide sequence SEQ ID NO:4234 - Homo sapiens, 833 aa. [WO200058473-A2, 05-OCT-2000]	1..274 42..315	270/274 (98%) 274/274 (99%)	e-150
ABB80078	Human kinesin motor protein (HsKrp5) amino acid sequence - Homo sapiens, 1279 aa. [US6379941-B1, 30-APR-2002]	1..274 488..761	271/274 (98%) 272/274 (98%)	e-149
AAM40604	Human polypeptide SEQ ID NO 5535 - Homo sapiens, 232 aa. [WO200153312-A1, 26-JUL-2001]	55..286 1..232	219/232 (94%) 226/232 (97%)	e-118
AAM38818	Human polypeptide SEQ ID NO 1963 - Homo sapiens, 229 aa. [WO200153312-A1, 26-JUL-2001]	64..286 7..229	218/223 (97%) 222/223 (98%)	e-118
AAY41675	Human channel-related molecule HCRM-3 - Homo sapiens, 229 aa. [WO9943807-A2, 02-SEP-1999]	64..286 7..229	218/223 (97%) 222/223 (98%)	e-118

5 In a BLAST search of public sequence databases, the NOV11a protein was found to have homology to the proteins shown in the BLASTP data in Table 11E.

Table 11E. Public BLASTP Results for NOV11a				
Protein Accession Number	Protein/Organism/Length	NOV11a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q9UF54	Hypothetical 96.7 kDa protein - Homo sapiens (Human). 833 aa	1..274 42..315	265/274 (96%) 269/274 (97%)	e-146

	(fragment).			
Q95LL1	Hypothetical 98.5 kDa protein - Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey), 865 aa (fragment).	1..256 610..865	245/256 (95%) 254/256 (98%)	e-135
Q95JP3	Hypothetical 49.3 kDa protein - Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey), 428 aa.	1..248 166..413	242/248 (97%) 247/248 (99%)	e-132
Q9QXL2	Kif21a - Mus musculus (Mouse), 1573 aa.	23..270 551..793	68/255 (26%) 129/255 (49%)	2e-16
Q64075	Nucleoporin p62 homolog protein - Rattus sp, 215 aa (fragment).	90..239 12..151	55/151 (36%) 86/151 (56%)	6e-13

PFam analysis predicts that the NOV11a protein contains the domains shown in the Table 11F.

Table 11F. Domain Analysis of NOV11a			
Pfam Domain	NOV11a Match Region	Identities/ Similarities for the Matched Region	Expect Value
No Significant Matches Found			

### Example 12.

5 The NOV12 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 12A.

Table 12A. NOV12 Sequence Analysis			
	SEQ ID NO: 143	2754 bp	
NOV12a, CG135823-01 DNA Sequence	ATTGCCCTGTAACCTGTCAAAGAAGAGCTAAGGGAGCTTTCGGGGTTGGCTTCTTGG AGGCTGCTTTCTCCTTTACTTGGAAGGCTTCGCTAGTGATGGACCCATACATGATTCA GATGAGCAGCAAAGGCAACCTCCCCTCAATTCTGGACGTGCATGTCAACGTTGGTGGG AGAAGCTCTGTGCCGGGAAAAATGAAAGGCAGAAAGGCCAGGTGGTCTGTGAGGCCCT CAGACATGGCCAAGAAAACCTTCAACCCCATCCGAGCCATTGTGGACAACATGAAGGT GAAACCAAATCCAAACAAACCATGATTCCCTGTCCATTGGGGACCCTACTGTGTTT GGAAACCTGCCTACAGACCCTGAAGTTACCCAGGCAATGAAAGATGCCCTGGACTCGG GCAAATATAATGGCTATGCCCATCCATCGGCTTCCTATCCAGTCGGGAGGAGATTGC TTCTTATTACCACTGTCTGAGGCACCCCTAGAAGCTAAGGACGTCATTCTGACAAGT GGCTGCAGCCAAGCTATTGACCTTTGTTTAGCTGTGTGGCCCAACCCAGGGCAGAACA TCCTGGTTCCAAGACCTGGTTTCTCTCTACAAGACTCTGGCTGAGTCTATGGGAAT TGAGGTCAAACCTACAATTTGTTGCCAGAGAAATCTGGGAAATTGACCTGAAACAA CTGGAATATCTAATTGATGAAAAGACAGCTTGTCTCATGTCAATAATCCATCAAACC CCTGTGGGTCAAGTGTTCAGCAAACGTCATCTTCAGAAGATTCTGGCAGTGGCTGCACG GCAGTGTGTCCCATCTTAGCTGATGAGATCTATGGAGACATGGTGTTCGGATTGC AAATATGAACCACTGGCCACCCTCAGCACCGATGTCCCATCCTGTCTGTGGAGGGC TGGCCAAGCGCTGGCTGGTTCTGGCTGGAGGTTGGGCTGGATCCTCATTATGACCG AAGAGACATTTTGGCAATGAGATCCGAGATGGGCTGGTGAAGCTGAGTCAGCGCATT		

	TTGGGACCTGTACCATTGTCCAGGGAGCTCTGAAAAGCATCCTATGTGCGACCCCGG GAGAGTTTACCACAACACTCTGAGCTTCCCAAGTCCAATGCTGATCTCTGTTATGG GCGTTGGCTGCCATCCCTGGACTCCGGCCAGTCCGCCCTCTGGGGCTATGTACCTC ATGGTTGGAATTGAGATGGAACATTTCCAGAATTTGAGAACGATGTGGAGTTCACGG AGCGGTTAGTTGCTGAGCAGTCTGTCCACTGCCTCCAGCAACGTGCTTTGAGTACCC GAATTCATCCGAGTGGTCATCACAGTCCCCGAGGTGATGATGCTGGAGGCGTGCAGC CGGATCCAGGAGTTCGTGAGCAGCACTACCATTGTGCTGAAGGCAGCCAGGAGGAGT GTGATAAATAGGCCCTGCATCCATTCTCCTGAGGATGTGTCCCATCTAGGGAAGGCTGG ACTAGGCCTTGCGGCTCCTCAGGGACTCAGGTGGCCCTACTGGGAGAGGGCCCTCAAA TGCACCATGTCAAGGGTTCAAGATTGTTCCCTGCTTTTCCCCAAGTACAACCACACCCA CACTCAGATCCTCCTCATTACATCGCAGATTACTCCCTTGCTCTGCGCTGCTAGAGT GACTCACTAATTCATTAATCTGCCTCCCTCTCGTAAGATTTCCTTCTTTTCTTG AAAGTACCAGGTGAACAAAGTTTACCAGAAAGCAGTTGAGACAAGAAAATAAGAGCTC AGGATGAGGGAAAAGAAAAGATTGAGAGAATTTGTGCCCCAACCATTTCTCAGAC TCTAAGAAAGAACACGCTCTCTCCAGGCAGGTCTGAAGCTCAACTCTCTTATGGCCTC ACTTCAGGTATACCTCACTTTACACAATAGAATTATAACTGGAAGAAGTTGGGGACA CATGTATTTGGTGATTACATTTTAAACACATTAGGAAAAGTTGCTATTTGAACTTTTT ATTGATTTTGGGGGAGTAAAGAATTATTTTGGATGCAATAAATATCTTTAATTG ATCGACTTGCCAAATTTAGATTTGTGTGCATCAGGCTTCTTTTCTTTTCTTTTCTAG AGAAGTTCAATATAAGCTTTTCTTTCTTTGTTCTTTCTTTCTTTTCTTTTCTTTTCTG AGTCTTGCTCTGTGCGCCATGCTGGAGTGCAGTGGCGCATCTCGGCTCACTGCAACC TCCACCTCCTGGGTTCAAGCGATTCTCTTGCTCAACCTCCCAAGCAGTTGGGACTAC AGGCGTGAGCCACCATGCCCGCTAATTTTGTATTTTGTAGAGACAGGTTTTCAC CATGTTAGCCAGGCTGCTCAAACCTCCTGACCTCAGGCAATCTGCCCGCTGGGCTCT CCTAAAGTACTGGGATTACAGGCGTGAGCCACCTCGCCAGCGGCATCAGGCTTTCTT AAAGTGAGAGCAGCCTGTACTAGAGCAAGCAGGAATCAGAGACCTTCCAGAAATACT ACTGTGTAAGGGCCAGAAATATCTCACTTGTCTATTGTTATATAATCATTATTACTTT TGCTGTAATGTTAATATTGATTTATTAATATATATATCTTTTCATACATTCTCTAAG AAACATTATATTGATAAGATCTTTTATTTTGCAGGGCATAAATTATTGTTTTTCTT TTTTTTTTTTAATAAATTTACCAAGT		
	ORF Start: ATG at 97	ORF Stop: TAG at 1459	
	SEQ ID NO: 144	454 aa	MW at 50398.8kD
NOV12a, CG135823-01 Protein Sequence	MDPYMIQMSKGNLPSILDVHVNVGGRSSVPGMKGRKARWSVRPSDMAKKTFFNP IIVDNMVKPNPNKTMISLSIGDPTVFGNLPTDPEVTQAMKDALDSGKYNGYAPSI GFLSSREEIASYYHCPEAPLEAKDVLTSGCSQAIDLCLAVLANPGQNLVPRPGFS LYKT LAESMGIEVKLYNLLPEKSWEIDLKQLEYLIDEKTA CLIVNPNPFCGSV FSKRHLQK ILAVALRQCVPILADETYGDMVFSCKYEPLATLSTDVPI LSCGGLAKRWLVPGWRLG WILIHDRRDIFGNEIRDGLVKLSQRILGPCTIVQ GALKSILCRTPGEFYHNTLSFLKS NADLCYALAAIPGLRPVRPSGAMYL MVGIEMEHFPEFENDVEFTERLVAEQSVHCLP ATCFEYPNFI RVVITVPEVMMLEACSRIQEFCEQHYHCAEGSQEEDCK		
	SEQ ID NO: 145	1400 bp	
NOV12b, CG135823-02 DNA Sequence	CCAGAATTCACCATGGACCCATACATGATTGAGATGAGCAGCAAGGCAACCTCCCC TCAATTCTGGACGTGCATGTCAACGTTGGTGGGAGAAGCTCTGTGCCGGGAAAAATGA AAGGCAGAAAGGCCAGGTGGTCTGTGAGGCCCTCAGACATGGCCAAGAAAACCTTCAA CCCCATCCGAGCCATTGTGGACAACATGAAGGTGAAACCAATCCAAACAAAACCATG ATTTCCCTGTCCATTGGGGACCTACTGTGTTTGGAAACCTGCCTACAGACCTGGAAG TTACCCAGGCAATGAAAGATGCCCTGGACTCGGGCAAATATAATGGCTATGCCCATC CATCGGCTTCTATCCAGTCGGGAGGAGATTGCTTCTTATTACCACTGTCTGAGGCA CCCCTAGAAGCTAAGGACGTCATTCTGACAAGTGGCTGCAGCCAAGCTATTGACCTTT GTTAGCTGTGTTGGCCAAACCCAGGGCAAAACATCCTGGTTCCAAGACCTGGTTTCTC TCTCTACAAGACTCTGGCTGAGTCTATGGGAATTGAGGTCAAACCTCTACAATTGTTG CCAGAGAAATCTGGGAAATTGACCTGAAACAACCTGGAATATCTAATTGATGAAAAGA CAGCTTGTCTCATTGTCAATAATCCATCAAACCCCTGTGGGTGAGTGTTCAGCAAACG TCATCTTCAGAAAGATTCTGGCAGTGGCTGCACGGCAGTGTGTCCCATCTTAGCTGAT GAGATCTATGGAGACATGTTGTTTTCGGATTGCAATATGAACCACTGGCCACCCCTCA GCACCGATGTCCCATCTGTCTGTGGAGGGCTGGCCAAGCGCTGGCTGGTTCTCTGG CTGGAGGTTGGGCTGGATCCTCATTATGACCGAAGAGACATTTTGGCAATGAGATC CGAGATGGGCTGGTGAAGCTGAGTCAGCGCATTTTGGGACCCGTGACCATTGTCCAGG		

	GAGCTCTGAAAAGCATCCTATGTGCGACCCCGGAGAGTTTTACCACAACACTCTGAGCTTCCTCAAGTCCAATGCTGATCTCTGTTATGGGGCGTTGGCTGCCATCCCTGGACTCCGGCCAGTCCGCCCTTCTGGGGCTATGTACCTCATGGTTGGAATTGAGATGGAACATTTCACAGAATTTGAGAACGATGTGGAGTTCACGGAGCGGTTAGTTGCTGAGCAGTCTGTCCACTGCCTCCCAGCAACGTGCTTTGAGTACCCGAATTTTCATCCGAGTGGTTCATCACAGTCCCCGAGGTGATGATGCTGGAGGCGTGCAGCCGGATCCAGGAGTTCTGTGAGCAGCCTACCATTGTGCTGAAGGCAGCCAGGAGGAGTGTGATAAATAGGGTGGCGGCCGCTTTTTCCTT		
	ORF Start: ATG at 14		ORF Stop: TAG at 1376
	SEQ ID NO: 146	454 aa	MW at 50398.8kD
NOV12b, CG135823-02 Protein Sequence	MDPYMIQMSSKGNLPSILDVHVNVGGRSSVPGMKGRKARWSVRPSDMAKKTFFNPIRAIVDNMVKVPNPNTMISLSIGDPTVFGNLPDPEVTQAMKDALDSGKYNGYAPSIGFLSSREEIASYYHCPEAPLEAKDVLTSGCSQAIDLCLAVLANPGQNILVPRPGFSLYKT LAESMGIEVKLYNLLPEKSWEIDLKQLEYLIDEKTACLI VNNPSNPCGSVFSKRHLQK ILAVAAARQCVPI LADEIY GDMVFS DCKYEPLATLSTDVPI LSCGGLAKRWLVPGWRLG WILIHDRRDI FGNEIRDGLVKLSQRILGPCTIVQGALKSILCRTPGEFYHNTLSFLKS NADLCY GALAAI PGLRPVRPSGAMYL MVGIEMEHFPEFENDVEFTERLVAEQSVHCLP ATCFEYPNFIRVVITVPEVMMLEACSRIQBFCEQHYHCAEGSQEECDK		
	SEQ ID NO: 147	1400 bp	
NOV12c, 233048273 DNA Sequence	CCAGAATTCACCATGGACCCATACATGATTGAGATGAGCAGCAAAGGCAACCTCCCC TCAATTCTGGACGTGCATGTCAACGTTGGTGGGAGAAGCTCTGTGCCGGGAAAAATGA AAGGCAGAAAGGCCAGGTGGTCTGTGAGGCCCTCAGACATGGCCAAGAAAACCTTCAA CCCCATCCGAGCCATTGTGGACAACATGAAGGTGAAACCAAATCCAAACAAAACCATG ATTTCCCTGTCCATTGGGGACCCTACTGTGTTTGGAAACCTGCCTACAGACCCCTGAAG TTACCCAGGCAATGAAAGATGCCCTGGACTCGGGCAAATATAATGGCTATGCCCATC CATCGGCTTCCTATCCAGTCGGGAGGAGATTGCTTCTTATTACCACTGTCTGAGGCA CCCCTAGAAGCTAAGGACGTCATTCTGACAAGTGGCTGCAGCCAAGCTATTGACCTTT GTTTAGCTGTGTTGGCCAACCCAGGGCAAAACATCCTGGTTCGAAGACCTGGTTTCTC TCTCTACAAGACTCTGGCTGAGTCTATGGGAATTGAGGTCAAACCTTACAATTTGTTG CCAGAGAAATCTTGGGAATTTGACCTGAAACAACCTGGAATATCTAATTGATGAAAAGA CAGCTTGTCTCATTTGTCAATAATCCATCAAACCCCTGTGGGTCACTGTTTCAGCAACG TCATCTTCAGAAGATTCTGGCAGTGGCTGCACGGCAGTGTGTCCCCATCTTAGCTGAT GAGATCTATGGAGACATGGTGTGTTTCGATTGCAAATATGAACCACTGGCCACCTCA GCACCGATGTCCCCATCCTGTCTGTGGAGGGCTGGCCAAGCGCTGGCTGGTTCTCTGG CTGGAGGTTGGGCTGGATCCTCATTCATGACCGAAGAGACATTTTGGCAATGAGATC CGAGATGGGCTGGTGAAGCTGAGTCAGCGCATTTTGGGACCTGTACATTGTGTCAGG GAGCTCTGAAAAGCATCCTATGTGCGACCCCGGAGAGTTTTACCACAACACTCTGAG CTTCTCAAGTCCAATGCTGATCTCTGTTATGGGGCGTTGGCTGCCATCCCTGGACTC CGGCCAGTCCGCCCTTCTGGGGCTATGTACCTCATGGTTGGAATTGAGATGGAACATT TCCCAGAATTTGAGAACGATGTGGAGTTCACGGAGCGGTTAGTTGCTGAGCAGTCTGT CCACTGCCTCCCAGCAACGTGCTTTGAGTACCCGAATTTTCATCCGAGTGGTGCATCAC GTCCCCGAGGTGATGATGCTGGAGGCGTGCAGCCGGATCCAGGAGTTCTGTGAGCAGC ACTACCATTGTGCTGAAGGCAGCCAGGAGGAGTGTGATAAATAGGGTGGCGGCCGCTT TTTTCCTT		
	ORF Start: at 2		ORF Stop: TAG at 1376
	SEQ ID NO: 148	458 aa	MW at 50829.2kD
NOV12c, 233048273 Protein Sequence	QNSTMDPYMIQMSSKGNLPSILDVHVNVGGRSSVPGMKGRKARWSVRPSDMAKKTFFNPIRAIVDNMVKVPNPNTMISLSIGDPTVFGNLPDPEVTQAMKDALDSGKYNGYAPS IGFLSSREEIASYYHCPEAPLEAKDVLTSGCSQAIDLCLAVLANPGQNILVPRPGFS LYKTLAESMGIEVKLYNLLPEKSWEIDLKQLEYLIDEKTACLI VNNPSNPCGSVFSKR HLQKILAVAAARQCVPI LADEIY GDMVFS DCKYEPLATLSTDVPI LSCGGLAKRWLVPG WRLGWILIHDRRDI FGNEIRDGLVKLSQRILGPCTIVQGALKSILCRTPGEFYHNTLS FLKSNADLCY GALAAI PGLRPVRPSGAMYL MVGIEMEHFPEFENDVEFTERLVAEQSV HCLPATCFEYPNFIRVVITVPEVMMLEACSRIQBFCEQHYHCAEGSQEECDK		
	SEQ ID NO: 149	1271 bp	
NOV12d,	CCAGAATTCACCATGGACCCATACATGATTGAGATGAGCAGCAAAGGCAACCTCCCC		

233048286 DNA Sequence	TCAATTCTGGACGTGCATGTCAACGTTGGTGGGAGAAGCTCTGTGCCGGGAAAAATGA AAGGCAGAAAGGCCAGGTGGTCTGTGAGGCCCTCAGACATGGCCAAGAAAACTTTCAA CCCCATCCGAGCCATTGTGGACAACATGAAGGTGAAACCAAATCCAAACAAAACCATG ATTTCCCTGTCCATTGGGGACCCTACTGTGTTTGGAAACCTGCCACAGACCCCTGAAG TTACCCAGGCAATGAAAGATGCCCTGGACTCGGGCAAATATAATGGCTATGCCCATC CATCGGCTTCTATCCAGTCGGGAGGAGATTGCTTCTTATTACCATGTCTCTGAGGCA CCCCTAGAAGCTAAGGACGTCATTCTGACAAGTGGCTGCAGCCAAGCTATTGACCTTT GTTTAGCTGTGTTGGCCAACCCAGGGCAAAACATCCTGGTTCCAAGACCTGGTTTCTC TCTCTACAAGACTCTGGCTGAGTCTATGGGAATTGAGGTCAAACCTCTACAATTGTGTG CCAGAGAAATCTTGGGAAATTGACCTGAAACAACCTGGAATATCTAATTGATGAAAAGA CAGCTTGTCTCATTGTCAATAATCCATCAAACCCCTGTGGGTGAGTGTTCAGCAAAACG TCATCTTCAGAAGATTCTGGCAGTGGCTGCACGGCAGTGTGTCCCCATCTTAGCTGAT GAGATCTATGGAGACATGGTGTTTTCGGATTGCAAATATGAACCACTGGCCACCCCTCA GCACCGATGTCCCCATCCTGTCTGTGGAGGGCTGGCCAAGCGCTGGCTGGTTCTCTGG CTGGAGGTTGGGCTGGATCCTCATTATGACCGAAGAGACATTTTGGCAATGAGTGTG AATGCTGATCTCTGTTATGGGGCGTTGGCTGCCATCCCTGGACTCGGCCAGTCCGCC CTTCTGGGGCTATGTACCTCATGGTTGGAATTGAGATGGAACATTTCCAGAAATTGGA GAACGATGTGGAGTTCACGGAGCGGTTAGTTGCTGAGCAGTCTGTCCACTGCCCTCCA GCAACGTGCTTTGAGTACCCGAATTTATCCGAGTGGTCATCAGTCCCCGAGGTGA TGATGCTGGAGGCGTGCAGCCGATCCAGGAGTTCTGTGAGCAGCACTACCATTGTGC TGAAGGCAGCCAGGAGGAGTGTGATAAATAGGGTGGCGGCCGCTTTTTCCTT		
	ORF Start: at 2		ORF Stop: TAG at 1247
	SEQ ID NO: 150	415 aa	MW at 46059.6kD
NOV12d, 233048286 Protein Sequence	QNSTMDPYMIQMSSKGNLPSILDVHVNVGGRSSVPGKMKGRKARWSVRPSDMAKKTFFN PIRAIVDNMKVKPNPNKTMISLSIGDPTVFGNLPDPEVTQAMKDALDSGKNGYAPS IGFLSSREEIASYYHCPEAPLEAKDVIITSGCSQAIDLCLAVLANPQNILVPRPGFS LYKTLAESMGIEVKLYNLLPEKSWEIDLKQLEYLIDKTAACLI VNNPSNPCGSVFSKR HLQKILAVAAARQCVPI LADEIYGDVFSCKYEPLATLSTDVPI LSCGGLAKRWLVPG WRLGWILIHRRDIFGNESNADLCYGALAAI PGLRPVRPSGAMVLMVGIEMEHFPEFE NDVEFTERLVAEQSVHCLPATCFEYPNFI RVVITVPEVMMLEACSRIQEFCEQHYHCA EGSQEECDK		
	SEQ ID NO: 151	1372 bp	
NOV12e, 248490358 DNA Sequence	ACCATGGACCCATACATGATTGAGATGAGCAGCAAAGGCAACCTCCCCTCAATTCTGG ACGTGCATGTCAACGTTGGTGGGAGAAGCTCTGTGCCGGGAAAAATGAAAGGCAGAAA GGCCAGGTGGTCTGTGAGGCCCTCAGACATGGCCAAGAAAACCTTCAACCCCATCCGA GCCATTGTGGACAACATGAAGGTGAAACCAAATCCAAACAAAACCATGATTTCCTGT CCATTGGGGACCCTACTGTGTTTGGAAACCTGCCTACAGACCTTGAAGTTACCCAGGC AATGAAAGATGCCCTGGACTCGGGCAAATATAATGGCTATGCCCCATCCATCGGCTTC CTATCCAGTCGGGAGGAGATTGCTTCTTATTACCACTGTCTGAGGCACCCCTAGAAG CTAAGGACGTCATTCTGACAAGTGGCTGCAGCCAAGCTATTGACCTTTGTTTAGCTGT GTTGGCCAACCCAGGGCAAAACATCCTGGTTCCAAGACCTGGTTCTCTCTCTACAAG ACTCTGGCTGAGTCTATGGGAATTGAGGTCAAACCTTACAATTGTTGCCAGAGAAAT CTTGGGAAATTGACCTGAAACAACCTGGAATATCTAATTGATGAAAAGACAGCTGTCT CATTGTCAATAATCCATCAAACCCCTGTGGGTGAGTGTTCAGCAAACGTCATCTTCAG AAGATTCTGGCAGTGGCTGCACGGCAGTGTGTCCCATCTTAGCTGATGAGATCTATG GAGACATGGTGTGTTTCGGATTGCAAATATGAACCACTGGCCACCCCTCAGCACCGATGT CCCCATCCTGTCTGTGGAGGGCTGGCCAAGCGCTGGCTGGTTCTGGCTGGAGGTTG GGCTGGATCCTCATTATGACCGAAGAGACATTTTGGCAATGAGATCCGAGATGGGC TGGTGAAGCTGAGTCAGCGCATTTTGGGACCCTGTACCATTGTCCAGGGAGCTCTGAA AAGCATCCTATGTGCGACCCCGGAGAGTTTACCACAACACTCTGAGCTTCTCAAG TCCAATGCTGATCTCTGTATGGGGCGTTGGCTGCCATCCCTGGACTCCGGCCAGTCC GCCCTTCTGGGGCTATGTACCTCATGGTTGGAATTGAGATGGAACATTTCCAGAAATT TGAGAACGATGTGGAGTTCACGGAGCGGTTAGTTGCTGAGCAGTCTGTCCACTGCCTC CCAGCAACGTGCTTTGAGTACCCGAATTTATCCGAGTGGTCATCAGTCCCCGAGG TGATGATGCTGGAGGCGTGCAGCCGATCCAGGAGTTCTGTGAGCAGCACTACCATTG TGCTGAAGGCAGCCAGGAGGAGTGTGATAAATAGGGTG		
	ORF Start: at 1		ORF Stop: TAG at 1366

	SEQ ID NO: 152	455 aa	MW at 50499.9kD
NOV12e, 248490358 Protein Sequence	TMDPYMIQMSSKGNLPSILDVHVNVGGRSSVPGMKGRKARWSVRPSDMAKKTFFNP AIVDNMKVKPNPNKTMISLSIGDPTVFGNLPTDPEVTQAMKDALDSGKYNGYAPSIGF LSSREEIASYYHCPEAPLEAKDVILTSGCSSQAIDLCLAVLANPGQNILVPRPGFSLYK TLAESMGIEVKLYNLLPEKSWEIDLKQLEYLIDEKTACLI VNNPSNPCGSVFSKRHLQ KILAVARQCVPILADEIYGDVMSDCKYEPLATLSTDVPILSCGGLAKRWLVPGWRL GWILIHRRDIFGNEIRDGLVKLSQRILGPCTIVQGALKSILCRTPGEFYHNTLSFLK SNADLCYALAAIPGLRPVRPSGAMVLMVGIEMEHFPEFENDVEFTEFLVAEQSVHCL PATCFEYPNFIRVVITVPEVMMLEACSRIQEFCEQHYHCAEGSQEECDK		
	SEQ ID NO: 153	1398 bp	
NOV12f, 254868693 DNA Sequence	ACCCATACATGATTGAGATGAGCAGCAAAGGCAACCTCCCTCAATTCTGGACGTGCA TGTCACCGTTGGTGGGAGAAGCTCTGTGCCGGGAAAAATGAAAGGCAGAAAGGCCAGG TGGTCTGTGAGGCCCTCAGACATGGCCAAGAAAACTTTCAACCCCATCCGAGCCATTG TGGACAACATGAAGGTGAAACCAATCCAACAAAACCATGATTTCCCTGTCCATTGG GGACCCTACTGTGTTTGGAAACCTGCCTACAGACCCTGAAGTTACCCAGGCAATGAAA GATGCCCTGGACTCGGGCAAATATAATGGCTATGCCCCATCCATCGGCTTCTATCCA GTCCGGGAGGAGATTGCTTCTTATTACCACTGTCTGAGGCACCCCTAGAAGCTAAGGA CGTCATTCTGACAAGTGGCTGCAGCCAAGCTATTGACCTTTGTTTAGCTGTGTTGGCC AACCAGGGCAAAACATCCTGGTTCCAAGACCTGGTTTCTCTCTCTACAAGACTCTGG CTGAGTCTATGGGAATTGAGGTCAAACCTCACAATTTGTTGCCAGAGAAATCTTGGGA AATTGACCTGAAACAACTGGAATATCTAATTGATGAAAAGACAGCTTGTCTCATTTGT AATAATCCATCAACCCCTGTGGGTGAGTGTTCAGCAAACGTCTCTCAGAAGATTC TGGCAGTGGCTGCAGGCAGTGTGTCCCATCTTAGCTGATGAGATCTATGGAGACAT GGTGTCTTCGGATTGCAAATATGAACCACTGGCCACCCCTCAGCACCGATGTCCCATC CTGTCTGTGGAGGGCTGGCCAAGCGCTGGCTGGTTCTGGCTGGAGGTTGGGCTGGA TCCTCATTCATGACCGAAGAGACATTTTGGCAATGAGATCCGAGATGGGCTGGTGAA GCTGAGTCAGCGCATTTTGGGACCCTGTACCATTTGTCCAGGGAGCTCTGAAAAGCATC CTATGTGCGACCCCGGAGAGTTTACCACAACACTCTGAGCTTCTCAAGTCCAATG CTGATCTCTGTTATGGGGCGTTGGCTGCCATCCCTGGACTCCGGCCAGTCCGCCCTTC TGGGGCTATGTACCTCATGGTTGGAATTGAGATGGAACATTTCCAGAATTTGAGAAC GATGTGGAGTTACCGAGCGGTTAGTTGCTGAGCAGTCTGTCCACTGCCAGTCCGACAA CGTGCTTTGAGTACCCGAATTCATCCGAGTGGTCATCACAGTCCCCGAGGTGATGAT GCTGGAGGCGTGAGCCGATCCAGGAGTTCTGTGAGCAGCACTACCATTGTGCTGAA GGCAGCCAGGAGGAGTGTGATAAATAGGGTGGCGGCCGCACTCGAGCACCACCACCAC CACCAC		
	ORF Start: at 3		ORF Stop: TAG at 1359
	SEQ ID NO: 154	452 aa	MW at 50152.5kD
NOV12f, 254868693 Protein Sequence	PYMIQMSSKGNLPSILDVHVNVGGRSSVPGMKGRKARWSVRPSDMAKKTFFNP DNMKVKPNPNKTMISLSIGDPTVFGNLPTDPEVTQAMKDALDSGKYNGYAPSIGFLSS REEIASYYHCPEAPLEAKDVILTSGCSSQAIDLCLAVLANPGQNILVPRPGFSLYKTLA ESMGIEVKLYNLLPEKSWEIDLKQLEYLIDEKTACLI VNNPSNPCGSVFSKRHLQKIL AVARQCVPILADEIYGDVMSDCKYEPLATLSTDVPILSCGGLAKRWLVPGWRLGWI LIHRRDIFGNEIRDGLVKLSQRILGPCTIVQGALKSILCRTPGEFYHNTLSFLKSNA DLCYALAAIPGLRPVRPSGAMVLMVGIEMEHFPEFENDVEFTEFLVAEQSVHCLPAT CFEYPNFIRVVITVPEVMMLEACSRIQEFCEQHYHCAEGSQEECDK		
	SEQ ID NO: 155	1414 bp	
NOV12g, 255667122 DNA Sequence	ACATCATCACCACCATCAGCACCACATACATGATTGAGATGAGCAGCAAAGGCAACCTC CCCTCAATTCTGGACGTGCATGTCAACGTTGGTGGGAGAAGCTCTGTGCCGGGAAAAA TGAAAGGCAGAAAGGCCAGGTGCTGTGAGGCCCTCAGACATGGCCAAGAAAACTTT CAACCCCATCCGAGCCATTGTGACAACATGAAGGTGAAACCAAAATCCAACAAAACC ATGATTTCCCTGTCCATTGGGGACCCTACTGTGTTTGGAAACCTGCCTACAGACCTG AAGTTACCCAGGCAATGAAAGATGCCCTGGACTCGGGCAAATATAATGGCTATGCCCC ATCCATCGGCTTCTATCCAGTCGGGAGGAGATTGCTTCTTATTACCACTGTCTGAG GCACCCCTAGAAGCTAAGGACGTATTCTGACAAGTGGCTGCAGCCAAGCTATTGACC TTTGTCTAGCTGTGTGGCCAACCCAGGGCAAAACATCCTGGTTCCAAGACCTGGTTT CTCTCTCTACAAGACTCTGGCTGAGTCTATGGGAATTGAGGTCAAACCTCACAATTTG TTGCCAGAGAAATCTGGGAAATTGACCTGAAACAACTGGAATATCTAATTGATGAAA		

	AGACAGCTTGTCTCATTGTCAATAATCCATCAAACCCCTGTGGGTGAGTGTTCAGCAA ACGTCATCTTCAGAAGATTCTGGCAGTGGCTGCACGGCAGTGTGTCCCATCTTAGCT GATGAGATCTATGGAGACATGGTGTTCGATTGCAAATATGAACCACTGGCCACCC TCAGCACCGATGTCCCATCTGTCTGTGGAGGGCTGGCCAAGCGCTGGCTGGTTCC TGGCTGGAGGTGGGTGGATCCTCATTCATGACCGAAGAGACATTTTGGCAATGAG ATCCGAGATGGGCTGGTGAAGCTGAGTCAGCGCATTTTGGGACCTGTACCATTGTCC AGGGAGCTCTGAAAAGCATCTATGTCGCACCCCGGAGAGTTTTACCACAACACTCT GAGCTTCCCTCAAGTCCAATGCTGATCTCTGTTATGGGGCGTTGGCTGCCATCCCTGGA CTCCGGCCAGTCCGCCCTTCTGGGGCTATGTACCTCATGGTTGGAATTGAGATGGAAC ATTTCCCAGAATTTGAGAACGATGTGGAGTTACGGAGCGGTTAGTTGCTGAGCAGTC TGTCCACTGCCTCCAGCAACGTGCTTTGAGTACCGAATTTTCATCCGAGTGGTCATC ACAGTCCCCGAGGTGATGATGCTGGAGGCGTGCAGCCGGATCCAGGAGTTCTGTGAGC AGCACTACCATTGTGCTGAAGGCAGCCAGGAGGTGTGATAAATAGGCGGCCGCACT CGAGCACCACCACCACCAC		
	ORF Start: at 2		ORF Stop: TAG at 1379
	SEQ ID NO: 156	459 aa	MW at 51090.4kD
NOV12g, 255667122 Protein Sequence	HHHHHHPYMIQMSSKGNLPSILDVHVNVGRRSSVPGMKGRKARWSVRPSDMAKKT NPRAIVDNMKVKPNPNKTMISLSIGDPTVFGNLPDPEVTQAMKDALDSGKYNFYAP SIGFLSSREEIASYYHCPEAPLEAKDVLTSQCSQAIDLCLAVLANPGQNILVPRPGF SLYKTLAESMGIEVKLYNLLPEKSWEIDLKQLEYLIDEKTACLI VNNPSNPGSVFVSK RHLQKILAVARQCVPI LADEIYGDVMSDCKYEPLATLSTDVPI LSCGGLAKRWLV P GWRLGWILIHRRDIFGNEIRDGLVKLSQRILGPCTIVQGALKSLCRTLPGEFYHNTL SFLKSNADLCYALAAI PGLRPVRPSGAMVLMVGIEMEHFPEFENDVEFTERLVAEQS VHCLPATCFEYPNFI RRVITVPEVMMLEACSRIQEFCEQHYHCAEGSQEBCDK		
	SEQ ID NO: 157	1412 bp	
NOV12h, 258252417 DNA Sequence	ACCCATACATGATTGAGATGAGCAGCAAAGGCAACCTCCCCTCAATCTGACGTCGA TGTCAACGTTGGTGGGAGAAGCTCTGTGCCGGGAAAAATGAAAGGCAGAAAGGCCAGG TGGTCTGTGAGGCCCTCAGACATGGCCAAGAAAACCTTCAACCCCATCCGAGCCATTG TGGACAACATGAAGGTGAACCAATCCAAACAAACCATGATTTCCTGTCCATTGG GGACCCTACTGTGTTTGGAAACCTGCCTACAGACCCTGAAGTTACCCAGGCAATGAAA GATGCCCTGGACTCGGGCAAATATAATGGCTATGCCCATCCATCGGCTTCTATCCA GTCCGGAGGAGATTGCTTCTTATTACCACTGTCTGAGGCACCCCTAGAAGCTAAGGA CGTCATTCTGACAAGTGGCTGCAGCCAAGCTATTGACCTTTGTTTAGCTGTGTTGGCC AACCAGGGGAAAACATCTGTTTCCAAGACCTGGTTCTCTCTTACAAGACTCTGG CTGAGTCTATGGGAATTGAGGTCAAACCTCTACAATTTGTTGCCAGAGAAATCTTGGGA AATTGACCTGAAACAACCTGGAATATCTAATTGATGAAAAGACAGCTGTCTCATTTGTC AATAATCCATCAAACCCCTGTGGGTGAGTTTCAGCAAACGTCATCTTCAAGAGATT TGGCAGTGGCTGCACGGCAGTGTGTCCCATCTTAGCTGATGAGATCTATGGAGACAT GGTGTTCGATTGCAAATATGAACCACTGGCCACCTCAGCAGCGATGTCCCATC CTGTCCTGTGGAGGGCTGGCCAAGCGCTGGCTGGTTCCTGGCTGGAGGTTGGGCTGGA TCCTCATTCTGACCGAAGAGACATTTTGGCAATGAGATCCGAGATGGGCTGGTGAA GCTGAGTCAGCGCATTTTGGGACCTGTACCATTGTCCAGGAGCTGTGAAAAGCATC CTATGTGCGACCCCGGGAGAGTTTTACCACAACACTCTGAGCTTCTCAAGTCCAATG CTGATCTCTGTTATGGGGCGTTGGCTGCCATCCCTGGACTCCGGCCAGTCCGCCCTTC TGGGGCTATGTACCTCATGTTGGAATTGAGATGGAACATTTCCAGAATTTGAGAAC GATGTGGAGTTCACGGAGCGGTTAGTTGCTGAGCAGTCTGTCCACTGCCTCCAGCAA CGTGCTTTGAGTACCCGAATTTCAATCCGAGTGGTCATCAGATCCCCGAGGTGATGAT GCTGGAGGCGTGCAGCCGATCCAGGAGTTCTGTGAGCAGCACTACCATTGTGCTGAA GGCAGCCAGGAGGAGTGTGATAAACATCATCACCACCATCACTAGGCGGCCGCACTCG AGCACCACCACCACCAC		
	ORF Start: at 3		ORF Stop: TAG at 1377
	SEQ ID NO: 158	458 aa	MW at 50975.4kD
NOV12h, 258252417 Protein Sequence	PYMIQMSSKGNLPSILDVHVNVGRRSSVPGMKGRKARWSVRPSDMAKKTNPRAIV DNMKVKPNPNKTMISLSIGDPTVFGNLPDPEVTQAMKDALDSGKYNFYAPSIGFLSS REEIASYYHCPEAPLEAKDVLTSQCSQAIDLCLAVLANPGQNILVPRPGFSLYKTLA ESMGIEVKLYNLLPEKSWEIDLKQLEYLIDEKTACLI VNNPSNPGSVFVSKRHLQKIL AVARQCVPI LADEIYGDVMSDCKYEPLATLSTDVPI LSCGGLAKRWLVPGWRLGWI		

	LIHRRDIFGNEIRDGLVKLSQRILGPCTIVQGALKSILCRTPGEFYHNTLSFLKSNA DLCYGALAAIPGLRPVRPSGAMLYLMVGIEMEHFPEFENDVEFTERLVAEQSVHCLPAT CFEYPNFIIRVVITVPEVMMLEACSRIQEFCEQHYHCAEGSQEEDKHHHHHH		
	SEQ ID NO: 159	1385 bp	
NOV12i, 259741773 DNA Sequence	CCATGGACCCATACATGATTGAGATGAGCAGCAAAGGCAACCTCCCCTCAATTCTGGGA CGTGCATGTCAACGTTGGTGGGAGAAGCTCTGTGCCGGGAAAAATGAAAGGCAGAAAG GCCAGGTGGTCTGTGAGGCCCTCAGACATGGCCAAGAAAACCTTTCAACCCCATCCGAG CCATTGTGGACAACATGAAGGTGAAACCAAATCCAAACAAAACCATGATTTCCCTGTC CATTGGGGACCCCTACTGTGTTTGGAAACCTGCCTACAGACCCTGAAGTTACCCAGGCA ATGAAAGATGCCCTGGACTCAGGCAAATATAATGGCTATGCCCCATCCATCGGCTTCC TATCCAGTCGGGAGGAGATTGCTTCTTATTACCACTGTCTGAGGCACCCCTAGAAGC TAAGGACGTCATTCTGACAAGTGGCTGCAGCCAAGCTATTGACCTTTGTTAGCTGTG TTGGCCAACCCAGGGCAAACATCCTGGTTCCAAGACCTGGTTTCTCTCTCTACAAGA CTCTGGCTGAGTCTATGGGAATTGAGGTCAAACCTCTACAATTGTTGCCAGAGAAATC TTGGGAAATTGACCTGAAACAACCTGGAATATCTAATTGATGAAAAGACAGCTTGTCTC ATTGTCAATAATCCATCAAACCCCTGTGGGTGAGTGTTCAGCAAACGTCATCTTCAGA AGATTCTGGCAGTGGCTGCACGGCAGTGTGTCCCCTCTTAGCTGATGAGATCTATGG AGACATGGTGTTTTCGATTGCAAATATGAACCACTGGCCACCCCTCAGCACCGATGTC CCCATCCTGTCTGTGGAGGGCTGGCCAAGCGCTGGCTGGTTCTGGCTGGAGGTTGG GCTGGATCCTCATTTCATGACCGAAGAGACATTTTGGCAATGAGATCCGAGATGGGCT GGTGAAGCTGAGTCAGCGCATTTTGGGACCCGTGACCATTGTCCAGGGAGCTCTGAAA AGCATCCTATGTGCGACCCCGGGAGAGTTTTACCACAACACTCTGAGCTTCTCAAGT CCAATGCTGATCTCTGTTATGGGGCGTTGGCTGCCATCCCTGGACTCCGGCCAGTCCG CCCTTCTGGGGCTATGTACCTCATGGTTGGAATTGAGATGGAACATTTCCCAAGATTT GAGAACGATGTGGAGTTACGGGAGCGGTTAGTTGCTGAGCAGTCTGTCCACTGCCCTCC CAGCAACGTGCTTTGAGTACCCGAATTTTCATCCGAGTGGTCATCAGTCCCCGAGGT GATGATGCTGGAGGCGTGCAGCCGGATCCAGGAGTTCTGTGAGCAGCACTACCATTGT GCTGAAGGCAGCCAGGAGGAGTGTGATAAACATCATCACCACCATCACTAG		
	ORF Start: ATG at 3		ORF Stop: TAG at 1383
	SEQ ID NO: 160	460 aa	MW at 51221.6kD
NOV12i, 259741773 Protein Sequence	MDPYMIQSSKGNLPSILDVHNVNVRSSVPGMKGRKARWSVRPSDMAKKTFFNPIRA IVDNMKVKPNPNKTMISLSIGDPTVFGNLPDPEVTQAMKDALDSGKNGYAPSIGFL SSREEIASYYHCPEAPLEAKDVILTSGCSQAIDLCLAVLANPGQNILVPRPGFSLYKT LAESMGIEVKLYNLLPEKSWEIDLKQLEYLIDEKTAELIVNPNPNCGSVFSKRHLQK ILAVARQCVPILADEIYGMVPSDKYEPLATLSTDVPIILSCGGLAKRWLVPGWRLG WILIHRRDIFGNEIRDGLVKLSQRILGPCTIVQGALKSILCRTPGEFYHNTLSFLKS NADLCYGALAAIPGLRPVRPSGAMLYLMVGIEMEHFPEFENDVEFTERLVAEQSVHCLP ATCFEYPNFIIRVVITVPEVMMLEACSRIQEFCEQHYHCAEGSQEEDKHHHHHH		
	SEQ ID NO: 161	1370 bp	
NOV12j, 260480043 DNA Sequence	CACCATGGACCCATACATGATTGAGATGAGCAGCAAAGGCAACCTCCCCTCAATTCTG GACGTGCATGTCAACGTTGGTGGGAGAAGCTCTGTGCCGGGAAAAATGAAAGGCAGAA AGGCCAGGTGGTCTGTGAGGCCCTCAGACATGGCCAAGAAAACCTTTCAACCCCATCCG AGCCATTGTGGACAACATGAAGGTGAAACCAAATCCAAACAAAACCATGATTTCCCTG TCCATTGGGGACCCCTACTGTGTTTGGAAACCTGCCTACAGACCCTGAAGTTACCCAGG CAATGAAAGATGCCCTGGACTCGGGCAAATATAATGGCTATGCCCCATCCATCGGCTT CCTATCCAGTCGGGAGGAGATTGCTTCTTATTACCACTGTCTGAGGCACCCCTAGAA GCTAAGGACGTCATTCTGACAAGTGGCTGCAGCCAAGCTATTGACCTTGTGTTAGCTG TGTTGGCCAACCCAGGGCAAACATCCTGGTTCCAAGACCTGGTTTCTCTCTCTACAA GACTCTGGCTGAGTCTATGGGAATTGAGGTCAAACCTCTACAATTGTTGCCAGAGAAA TCTTGGGAAATTGACCTGAAACAACCTGGAATATCTAATTGATGAAAAGACAGCTTGT TCATTGTCAATAATCCATCAAACCCCTGTGGGTGAGTGTTCAGCAAACGTCATCTTCA GAAGATTCTGGCAGTGGCTGCACGGCAGTGTGTCCCCTCTTAGCTGATGAGATCTAT GGAGACATGGTGTTTTCGATTGCAAATATGAACCACTGGCCACCCCTCAGCACCGATG TCCCCATCCTGTCTGTGGAGGGCTGGCCAAGCGCTGGCTGGTTCTGGCTGGAGGTT GGGCTGGATCCTCATTCATGACCGAAGAGACATTTTGGCAATGAGATCCGAGATGGG CTGGTGAAGCTGAGTCAGCGCATTTTGGGACCCGTGACCATTGTCCAGGGAGCTCTGA AAAGCATCCTATGTGCGACCCCGGAGAGTTTACCACAACACTCTGAGCTTCTCTCAA		

	GTCCAATGCTGATCTCTGTTATGGGGCGTTGGCTGCCATCCCTGGACTCCGGCCAGTC CGCCCTTCTGGGGCTATGTACCTCATGGTTGGAATTGAGATGGAACATTCCCAGAAT TTGAGAACGATGTGGAGTTCACGGAGCGGTTAGTTGCTGAGCAGTCTGTCCACTGCCT CCCAGCAACGTGCTTTGAGTACCCGAATTCATCCGAGTGGTCATCAGTCCCCGAG GTGATGATGCTGGAGGCGTGACGCCGATCCAGGAGTTCTGTGAGCAGCACTACCATT GTGCTGAAGGCAGCCAGGAGGAGTGTGATAAATAGG		
	ORF Start: at 2		ORF Stop: TAG at 1367
	SEQ ID NO: 162	455 aa	MW at 50499.9kD
NOV12j, 260480043 Protein Sequence	TMDPYMIQMSSKGNLPSILDVHVNVGGRSSVPGKMKGRKARWSVRPSDMAKKTFFNPIR AIVDNMVKPNPNKTMISLSIGDPTVFGNLPDPEVTQAMKDALDSGKNGYAPSIGF LSSREEIASYYHCPEAPLEAKDVILTSGCSQAIDLCLAVLANPGQNILVPRPGFSLYK TLAESMGIEVKLYNLLPEKSWEIDLKQLEYLIDEKTACLI VNNPSNPGSVFSKRHLQ KILAVARQCVPI LADEIYGD MVFSDCKYEPLATLSTDVPI LSCGGLAKRWLVPGWRL GWILIHDRRDI FGNEIRDGLVKLSQRILGPCTIVQGALKSILCRTPGEFYHNTLSFLK SNADLCYALAAI PGLRPVRPSGAMYLMVGIEMEHFP EFENDVEFTERLVAEQSVHCL PATCFEYPNFIRVITVPEVMMLEACSRIQEFCEQHYHCAEGSQEEDCK		
	SEQ ID NO: 163	1414 bp	
NOV12k, CG135823-03 DNA Sequence	ACATCATCACCACCATCACGACCATACATGATTCAGATGAGCAGCAAAGGCAACCTC CCCTCAATTCTGGACGTGCATGTCAACGTTGGTGGGAGAAGCTCTGTGCCGGGAAAAA TGAAAGGCAGAAAGGCCAGGTGGTCTGTGAGGCCCTCAGACATGGCCAAAGAAAACCTT CAACCCCATCCGAGCCATTGTGGACAACATGAAGGTGAAACCAAATCCAAACAAAACC ATGATTTCCCTGTCCATTGGGGACCTACTGTGTTTGGAAACCTGCCTACAGACCCTG AAGTTACCCAGGCAATGAAAGATGCCCTGGACTCGGGCAAATATAATGGCTATGCCCC ATCCATCGGCTTCCTATCCAGTCGGGAGGAGATTGCTTCTTATTACCACTGTCTCTGAG GCACCCCTAGAAGCTAAGGACGTCAATTCTGACAAGTGGCTGCAGCCAGCTATTGACC TTTGTGTTAGCTGTGTTGGCCAACCCAGGGCAAAACATCCTGGTTCCAAGACCTGGTTT CTCTCTCTACAAGACTCTGGCTGAGTCTATGGGAATTGAGGTCAAACCTCTACAATTTG TTGCCAGAGAAATCTGGGAAATTGACCTGAAACAACTGGAATATCTAATTGATGAAA AGACAGCTTGTCTCATTGTCAATAATCCATCAAACCCCTGTGGGTCACTGTTTCAGCAA ACGTCATCTTCAGAAGATTCTGGCAGTGGCTGCACGGCAGTGTGTCCCCATCTTAGCT GATGAGATCTATGGAGACATGGTGTGTTTCGGATTGCAAAATGAACCACTGCCACCC TCAGCACCGATGTCCCCATCCTGTCTGTGGAGGGCTGGCCAAGCGCTGGCTGGTTCC TGGCTGGAGGTTGGGCTGGATCCTCATTGATGACCGAAGAGACATTTTGGCAATGAG ATCCGAGATGGGCTGGTGAAGCTGAGTCAGCGCATTTTGGGACCTGTACCATTGTCC AGGGAGCTCTGAAAAGCATCCTATGTGCGACCCCGGGAGAGTTTACCACAACACTCT GAGCTTCTCAAGTCCAATGCTGATCTCTGTATGAGGGCGTTGGCTGCCATCCCTGGA CTCCGGCCAGTCCGCCCTTCTGGGGCTATGTACCTCATGGTTGGAATTGAGATGGAAC ATTTCCAGAAATTTGAGAACGATGTGGAGTTCACGGAGCGGTAGTTGCTGAGCAGTC TGTCCACTGCCTCCAGCAACGTGCTTTGAGTACCCGAATTCATCCGAGTGGTCATC ACAGTCCCCGAGGTGATGATGCTGGAGGCGTGACGCCGATCCAGGAGTTCTGTGAGC AGCACTACCATTGTGCTGAAGGCAGCCAGGAGGAGTGTGATAAATAGGCGGCCGCACT CGAGCACCACCACCACCAC		
	ORF Start: at 2		ORF Stop: TAG at 1379
	SEQ ID NO: 164	459 aa	MW at 51090.4kD
NOV12k, CG135823-03 Protein Sequence	HHHHHDPYMIQMSSKGNLPSILDVHVNVGGRSSVPGKMKGRKARWSVRPSDMAKKTFF NPIRAIVDNMVKPNPNKTMISLSIGDPTVFGNLPDPEVTQAMKDALDSGKNGYAP SIGFLSSREEIASYYHCPEAPLEAKDVILTSGCSQAIDLCLAVLANPGQNILVPRPGF SLYKTLAESMGIEVKLYNLLPEKSWEIDLKQLEYLIDEKTACLI VNNPSNPGSVFSK RHLQKILAVARQCVPI LADEIYGD MVFSDCKYEPLATLSTDVPI LSCGGLAKRWLV GWRLGWILIHDRRDI FGNEIRDGLVKLSQRILGPCTIVQGALKSILCRTPGEFYHNTL SFLKSNADLCYALAAI PGLRPVRPSGAMYLMVGIEMEHFP EFENDVEFTERLVAEQS VHCLPATCFEYPNFIRVITVPEVMMLEACSRIQEFCEQHYHCAEGSQEEDCK		
	SEQ ID NO: 165	1412 bp	
NOV12l, CG135823-04 DNA Sequence	ACCCATACATGATTCAGATGAGCAGCAAAGGCAACCTCCCTCAATTCTGGACGTGCA TGTCAACGTTGGTGGGAGAAGCTCTGTGCCGGGAAAAATGAAAGGCAGAAAGGCCAGG TGGTCTGTGAGGCCCTCAGACATGGCCAAGAAAACCTTCAACCCCATCCGAGCCATTG		

DNA Sequence	TGGACAACATGAAGGTGAAACCAAATCCAAACAAAACCATGATTTCCTGTCCATTGG GGACCCTACTGTGTTTGGAAACCTGCCTACAGACCCTGAAGTTACCCAGGCAATGAAA GATGCCCTGGACTCGGGCAAATATAATGGCTATGCCCCATCCATCGGCTTCCTATCCA GTCGGGAGGAGATTGCTTCTTATTACCACTGTCTGTAGGCACCCCTAGAAGCTAAGGA CGTCATTCTGACAAGTGGCTGCAGCCAAGCTATTGACCTTTGTTTAGCTGTGTTGGCC AAGCCAGGGCAAACATCCTGGTTCCAAGACCTGGTTTCTCTCTACAAGACTCTGG CTGAGTCTATGGGAATTGAGGTCAAACCTCTACAATTTGTTGCCAGAGAAATCTTGGGA AATTGACCTGAAACAACCTGGAATATCTAATTGATGAAAAGACAGCTTGTCTCATGTCT AATAATCCATCAAACCCCTGTGGGTCAAGTGTTCAGCAAACGTCATCTTCAGAAGATTCT TGGCAGTGGCTGCACGGCAGTGTGTCCCCATCTTAGCTGATGAGATCTATGGAGACAT GGTGTTTTTCGGATTGCAAATATGAACCACTGGCCACCCTCAGCACCGATGTCCCCATC CTGTCTGTGGAGGGCTGGCCAAGCGCTGGCTGGTTTCTGGCTGGAGGTTGGGCTGGA TCCTCATTCATGACCGAAGAGACATTTTTGGCAATGAGATCCGAGATGGGCTGGTGAA GCTGAGTCAGCGCATTTTGGGACCCGTGACCATTTGTCCAGGGAGCTCTGAAAAGCATC CTATGTGCGACCCCGGAGAGTTTTACCACAACACTCTGAGCTTCTCTCAAGTCCAATG CTGATCTCTGTTATGGGGCGTTGGCTGCCATCCCTGGACTCCGGCCAGTCCGCCCTTC TGGGGCTATGTACCTCATGGTTGGAATTGAGATGGAACATTTCCAGAAATTTGAGAAC GATGTGGAGTTCACGGAGCGGTTAGTTGCTGAGCAGTCTGTCCACTGCCTCCAGCAA CGTGCTTTGAGTACCCGAATTTTCATCCGAGTGGTCATCACAGTCCCCGAGGTGATGAT GCTGGAGGCGTGCAGCCGATCCAGGAGTTCTGTGAGCAGCACTACCATTGTGCTGAA GGCAGCCAGGAGGAGTGTGATAAACATCATCACCACCATCACTAGGCGGCCGCACTCG AGCACCACCACCACCACCAC		
	ORF Start: ATG at 9		ORF Stop: TAG at 1377
	SEQ ID NO: 166	456 aa	MW. at 50715.1kD
NOV12l, CG135823-04 Protein Sequence	MIQMSSKGNLPSILDVHVNVGGRSSVPGKMKGRKARWSVRPSDMAKKTFFNPRAIVDN MKVKPNPNKMTISLSIGDPTVFGNLPDPEVTQAMKDALDSGKNGYAPSIGFLSSRE ETIASYYHCPEAPLEAKDVLTSQCSQAIDLCLAVLANPGQNILVPRPGFSLYKTLAES MGEIVKLYNLLPEKSWEIDLKQLEYLIDEKTACLI VNNPSNPGCSVFSKRHLQKILAV AARQCVPILADEIYGDVMFSDCKYEPLATLSTDVPI LSCGGLAKRWLVPGWRLGWILI HRRRDI FGNEIRDGLVKLSQRILGPCTIVQGALKSILCRTPGEFYHNTLSFLKSNADL CYGALAAI PGLRPVRPSGAMYL MVGIEMEHFPEFENDVEFTERLVAEQSVHCLPATCF EYPNFI RRVVITVPEVMMLEACSRIQEFCEQHYHCAEGSQEECDKHHHHHH		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 12B.

Table 12B. Comparison of NOV12a against NOV12b through NOV12l.		
Protein Sequence	NOV12a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV12b	1..454 1..454	454/454 (100%) 454/454 (100%)
NOV12c	1..454 5..458	454/454 (100%) 454/454 (100%)
NOV12d	1..454 5..415	411/454 (90%) 411/454 (90%)
NOV12e	1..454 2..455	454/454 (100%) 454/454 (100%)
NOV12f	3..454 1..452	452/452 (100%) 452/452 (100%)
NOV12g	2..454	453/453 (100%)

	7..459	453/453 (100%)
NOV12h	3..454 1..452	452/452 (100%) 452/452 (100%)
NOV12i	1..454 1..454	454/454 (100%) 454/454 (100%)
NOV12j	1..454 2..455	454/454 (100%) 454/454 (100%)
NOV12k	2..454 7..459	453/453 (100%) 453/453 (100%)
NOV12l	5..454 1..450	450/450 (100%) 450/450 (100%)

Further analysis of the NOV12a protein yielded the following properties shown in Table 12C.

Table 12C. Protein Sequence Properties NOV12a	
PSort analysis:	0.6500 probability located in cytoplasm; 0.1000 probability located in mitochondrial matrix space; 0.1000 probability located in lysosome (lumen); 0.0000 probability located in endoplasmic reticulum (membrane)
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV12a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 12D.

Table 12D. Geneseq Results for NOV12a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV12a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
ABB58136	Drosophila melanogaster polypeptide SEQ ID NO 1200 - Drosophila melanogaster, 501 aa. [WO200171042-A2, 27-SEP-2001]	37..442 75..481	212/411 (51%) 296/411 (71%)	e-128
AAG10932	Arabidopsis thaliana protein fragment SEQ ID NO: 9454 - Arabidopsis thaliana, 407 aa. [EP1033405-A2, 06-SEP-2000]	68..441 8..385	136/382 (35%) 220/382 (56%)	3e-67
AAG10931	Arabidopsis thaliana protein fragment SEQ ID NO: 9453 - Arabidopsis thaliana, 445 aa.	68..441 46..423	136/382 (35%) 220/382 (56%)	3e-67

	[EP1033405-A2, 06-SEP-2000]			
AAG10930	Arabidopsis thaliana protein fragment SEQ ID NO: 9452 - Arabidopsis thaliana, 466 aa. [EP1033405-A2, 06-SEP-2000]	68..441 67..444	136/382 (35%) 220/382 (56%)	3e-67
AAG39068	Arabidopsis thaliana protein fragment SEQ ID NO: 48288 - Arabidopsis thaliana, 407 aa. [EP1033405-A2, 06-SEP-2000]	68..441 8..385	135/382 (35%) 219/382 (56%)	3e-66

In a BLAST search of public sequence databases, the NOV12a protein was found to have homology to the proteins shown in the BLASTP data in Table 12E.

Table 12E. Public BLASTP Results for NOV12a				
Protein Accession Number	Protein/Organism/Length	NOV12a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
P17735	Tyrosine aminotransferase (EC 2.6.1.5) (L-tyrosine:2-oxoglutarate aminotransferase) (TAT) - Homo sapiens (Human), 454 aa.	1..454 1..454	454/454 (100%) 454/454 (100%)	0.0
Q8QZR1	Similar to tyrosine aminotransferase (Hypothetical 50.6 kDa protein) - Mus musculus (Mouse), 454 aa.	1..454 1..454	418/454 (92%) 439/454 (96%)	0.0
P04694	Tyrosine aminotransferase (EC 2.6.1.5) (L-tyrosine:2-oxoglutarate aminotransferase) (TAT) - Rattus norvegicus (Rat), 454 aa.	1..454 1..454	416/454 (91%) 436/454 (95%)	0.0
Q9XSW4	Tyrosine aminotransferase - Mustela vison (American mink), 454 aa.	1..454 1..454	417/454 (91%) 438/454 (95%)	0.0
Q9QWS4	Tyrosine aminotransferase - Rattus norvegicus (Rat), 454 aa.	1..454 1..454	415/454 (91%) 435/454 (95%)	0.0

PFam analysis predicts that the NOV12a protein contains the domains shown in the Table 12F.

Table 12F. Domain Analysis of NOV12a			
Pfam Domain	NOV12a Match Region	Identities/ Similarities for the Matched Region	Expect Value

aminotran_1_2	113..438	72/356 (20%) 262/356 (74%)	2.1e-76
---------------	----------	-------------------------------	---------

**Example 13.**

The NOV13 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 13A.

Table 13A. NOV13 Sequence Analysis			
	SEQ ID NO: 167	1894 bp	
NOV13a, CG140122-01. DNA Sequence	CGCCGCTCGCCGACAGCTTACTTCCCGGCTCAGCAGGGAAGGTTCTAGAAAGTGA GCGCGGACGGTATGCAAAGTTGTGAATCCAGTGGTGACAGTGCGGATGACCCCTCTCAG TCGCGGCCTACGGAGAAGGGGACAGCCTCGTGTGGTGGTGATCGGCGCCGGCTTGGCT GGCCTGGCTGCAGCCAAAGCACTTCTTGAGCAGGGTTTCACGGATGTCACTGTGCTTG AGGCTTCCAGCCACATCGGAGGCCGTGTGCAGAGTGTGAAACTTGGACACGCCACCTT TGAGCTGGGAGCCACCTGGATCCATGGCTCCCATGGGAACCTATCTATCATCTAGCA GAAGCCAACGGCCTCCTGGAAGAGACAACCGATGGGGAACGCAGCGTGGGCCGCATCA GCCTCTATTCCAAGAATGGCGTGGCTGCTACCTTACCAACCACGGCCGAGGATCCC CAAGGACGTGGTTGAGGAATTGAGCATTATACAACGAGGTCTATAACTTGACCCAG GAGTTCTTCCGGCACGATAAACAGTCAATGCTGAAAGTCAAATAGCGTGGGGGTGT TCACCCGAGAGGAGGTGCGTAACCGCATCAGGAATGACCTGACGACCCAGAGGCTAC CAAGCGCCTGAAGCTCGCCATGATCCAGCAGTACCTGAAGTGGAGAGCTGTGAGAGC AGCTCACACAGCATGGACGAGGTGTCCCTGAGCGCCTTCGGGGAGTGGACCGAGATCC CCGGCGCTCACCACATCATCCCTCGGGCTTCATGCGGGTTGTGGAGCTGTGGCGGA GGGCATCCCTGCCACGTCATCCAGCTAGGGAACCTGTCCGCTGCATTCACTGGGAC CAGGCCTCAGCCCGCCAGAGGCCCTGAGATTGAGCCCCGGGGTGAGGGCGACACCA ATCAGGACTGAGGAGGGTGGCCAGGGTGGAGAGGAGCCCCGGGGGGCAGGTGGGA TGAGGATGAGCAGTGGTGGTGGTGGTGGAGTGCAGGACCGTGAGCTGATCCCGGCG GACCATGTGATTGTGACCGTGTGCTAGGTGTGCTAAAGAGGCAGTACACAGTTTCT TCCGGCCAGGCCTGCCACAGAGAAGGTGGCTGCCATCCACCGCCTGGGCATTGGCAC CACGACAAGATCTTCTGGAATTCGAGGAGCCCTTCTGGGGCCCTGAGTGCAACAGC CTACAGTTTGTGTGGAGGACGAAGCGGAGAGCCACACCTCACCTACCCACCTGAGC TCTGGTACCGCAAGATCTGCGGCTTTGATGTCTCTACCCGCCTGAGCGCTACGGCCA TGTGCTGAGCGGCTGGATCTGCGGGAGGAGGCCCTCGTCATGGAGAAGTGTGATGAC GAGGCAGTGGCCGAGATCTGCACGGAGATGCTGCGTCAGTTACAGGGAACCCCAACA TTCCAAACCTCGGCGAATCTTGCCTCGGCCTGGGGCAGCAACCTTACTTCCGTGG CTCCTATTCATACACGAGGTGGCTCCAGCGGGCGGATGTGGAGAAGCTGGCCAAG CCCTGCGGTACACGGAGAGCTCAAAGACAGCGCCATGCAGGTGCTGTTTCCGTG AGGCCACCCACCGAAGTACTATTCCACCACCCACGGTGTCTGCTGTCGGCCAGCG TGAGGCTGCCCGCCTCATTGAGATGTACCGAGACCTCTCCAGCAGGGGACCTGAGGG CTGTCTCGCTGCTGAGAAGAGCCACTAACTCGTGACCTCCAGCCTGCCCTTGTCTGC CGTGTGCTCTGCCTTCTGATCCTCTGTAGAAAGGATTTTATCTTCTGTAGAGCTA GCCGCCCTGACTGCCTTCAGACCTGGCCCTGTAGCTTT		
	ORF Start: ATG at 70		ORF Stop: TGA at 1735
	SEQ ID NO: 168	555 aa	MW at 61871.7kD
NOV13a, CG140122-01 Protein Sequence	MQSCSSGDSADDPLSRGLRRRQPRVVVIGAGLAGLAAAKALLEQGFTDVTVLEASS HIGGRVQSVKLGHATFELGATWIHSGHNPIYHLAEANGLLEETTDGERSVGRISLYS KNGVACYLTNHGRRIPKDVVEEFSDLYNEVYNLTQEFFRHKDPVNAESQNSVGVFTR EVRNRIRNDPDDPEATKRLKAMIQYLYKVESCESSSHSMDEVSLSAFGEWTEIPGAH HTIPSGFMRVVELLAEGIPAHVIQLGKPVRCIHWDAQASARPRGPEIEPRGEGDHNHDT GEGGQGGEEPGRGWEDEQWSVVVECEDRELIPADHVIIVTVSLGVLKQYTSFPRPG LPTEKVAAIHRLGIGTTDKIFLEFEEPFWGPENCNSLQFVWEDEAESHTLTYPPELWYR KICGFDVLYPPERYGHVLSGWI CGEEALVMEKCDDEAVAEICTEMLRQFTGNPNIPKP RRILRSAGWSNPYFRGSYSYTVGSSGADVEKLAKEPLPYTESSKTAPMQVLFSGEATH RKYYSTTHGALLSGQREARLIEMYRDLFQQGT		

	SEQ ID NO: 169	1012 bp	
NOV13b, 246864043 DNA Sequence	CACCATGCAAAGTTGTGAATCCAGTGGTGACAGTGCGGATGACCCCTCTCAGTCGCGGC CTACGGAGAAGGGGACAGCCTCGTGTGGTGGTGATCGGCGCCGGCTTGGCTGGCCTGG CTGCAGCCAAAGCACTTCTTGAGCAGGGTTTCACGGATGTCACTGTGCTTGAGGCTTC CAGCCACATCGGAGGCCGTGTGCAGAGTGTGAAACTTGGACACGCCACCTTTGAGCTG GGAGCCACCTGGATCCATGGCTCCCATGGGAACCCATCTATCATCTAGCAGAAGCCA ACGGCCTCCTGGAAGAGACAACCGATGGGAACGCAGCGTGGGCCGCATCAGCCTCTA TTCCAAGAATGGCGTGGCCTGCTACCTTACCAACCACGGCCGAGGATCCCCAAGGAC GTGGTTGAGGAATTCAGCGATTTATACAACGAGGTCTATAACTTGACCCAGGAGTTCT TCCGGCACGATAAACCAGTCAATGCTGAAAGTCAAAATAGCGTGGGGGTGTTACCCG AGAGGAGGTGCGTAACCGCATCAGGAATGACCCTGACGACCCAGAGGCTACCAAGCGC CTGAAGCTCGCCATGATCCAGCAGTACCTGAAGGTGGAGAGCTGTGAGAGCAGCTCAC ACAGCATGGACGAGGTGTCCCTGAGCGCCTTCGGGGAGTGGACCCGAGATCCCCGGCGC TCACCACATCATCCCCCTCGGGCTTCATGCGGGTTGTGGAGCTGCTGGCGGAGGGCATC CCTGCCCACGTATCCAGTAGGGAAACCTGTCCGCTGCATTCACTGGGACCAGGCCCT CAGCCCGCCCCAGAGGCCCTGAGATTGAGCCCCCTGCCGTACACAGAGAGCTCAAAGAC AGCGCCCATGCAGGTGCTGTTTTCCGGTGAGGCCACCCACCGCAAGTACTATTCCACC ACCACGGTGCTGTGCTGTCCGGCCAGCGTGAGGCTGCCCGCCTCATTGAGATGTACC GAGACCTCTTCAGCAGGGGACCTGA		
	ORF Start: at 2		ORF Stop: TGA at 1010
	SEQ ID NO: 170	336 aa	MW at 37093.2kD
NOV13b, 246864043 Protein Sequence	TMQSCSSGDSADDPLSRGLRRRQPRVVVIGAGLAGLAAAKALLEQFTDVTVLEAS SHIGGRVQSVKLGHATFELGATWIHSGHNP IYHLAEANGLLEETDGERSVGRISLY SKNGVACYLTNHGRRIPKDVVEEFSDLYNEVYNLTQEFFRHDKPVNAESQNSVGVFTR EEVRNRIRNDPDDPEATKRLKLAMIQQYLKVESCESSSHSMDEVSLSAFGEWTEIPGA HHII PSGFMRVVELLAEGIPAHVIQLKPVRCIHWDQASARPRGPEIEPLPYTESSKT APMQVLFSGEATHRKYYSTTHGALLSGORBAARLIEMYRDLFQQGT		
	SEQ ID NO: 171	1603 bp	
NOV13c, 246864086 DNA Sequence	CACCATGCAAAGTTGTGAATCCAGTGGTGACAGTGCGGATGACCCCTCTCAGTCGCGGC CTACGGAGAAGGGGACAGCCTCGTGTGGTGGTGATCGGCGCCGGCTTGGCTGGCCTGG CTGCAGCCAAAGCACTTCTTGAGCAGGGTTTCACGGATGTCACTGTGCTTGAGGCTTC CAGCCACATCGGAGGCCGTGTGCAGAGTGTGAAACTTGGACACGCCACCTTTGAGCTG GGAGCCACCTGGATCCATGGCTCCCATGGGAACCCATCTATCATCAGTACGAGAAGCCA ACGGCCTCCTGGAAGAGACAACCGATGGGAACGCAGCGTGGGCCGCATCAGCCTCTA TTCCAAGAATGGCGTGGCCTGCTACCTTACCAACCACGGCCGAGGATCCCCAAGGAC GTGGTTGAGGAATTCAGCGATTTATACAACGAGGTCTATAACTTGACCCAGGAGTTCT TCCGGCACGATAAACCAGTCAATGCTGAAAGTCAAAATAGCGTGGGGGTGTTACCCG AGAGGAGGTGCGTAACCGCATCAGGAATGACCCTGACGACCCAGAGGCTACCAAGCGC CTGAAGCTCGCCATGATCCAGCAGTACCTGAAGGTGGAGAGCTGTGAGAGCAGCTCAC ACAGCATGGACGAGGTGTCCCTGAGCGCCTTCGGGGAGTGGACCGAGATCCCCGGCGC TCACCACATCATCCCCCTCGGGCTTCATGCGGGTTGTGGAGCTGCTGGCGGAGGGCATC CCTGCCCACGTATCCAGCTAGGGAAACCTGTCCGCTGCATTCACTGGGACCAGGCCCT CAGCCCGCCCCAGAGGCCCTGAGATTGAGCCCCGGGTGTGCTAAAGAGGCAGTACAC CAGTTTCTTCGGCCAGGCCTGCCACAGAGAAGGTGGCTGCCATCCACCGCCTGGGC ATTGGCACCAACGACAAGATCTTTCTGGAATTCGAGGAGCCCTTCTGGGGCCCTGAGT GCAACAGCCTACAGTTTGTGTGGGAGGACGAAGCGGAGAGCCACACCCTCACCTACCC ACCTGAGCTCTGGTACCACAAGATCTGCGGCTTTGATGTCTCTACCCGCCTGAGCGC TACGGCCATGTGCTGAGCGGCTGGATCTGCGGGGAGGAGGCCCTCGTCATGGAGAAGT GTGATGACGAGGCAGTGGCCGAGATCTGCACGGAGATGCTGCTGAGTTCACAGGGAA CCCCACATTCCAAAACCTCGGCGAATCTTGCCTCGGCCTGGGCGACGAACCCCTTAC TTCCGCGGCTCCTATTATACACGCAGGTGGGCTCCAGCGGGGCGGATGTGGAGAAGC TGGCCAAGCCCTGCCGTACACGGAGAGCTCAAAGACAGCGCATGGAAGCTCCACAAA GCAGCAGCCTGGTCACCTTTTCTCTTCCAAGTGCCAGAACAGCCCTGGATGCTAAC AGGGGCGCCGTAAAGCCCATGCAGGTGCTGTTTTCCGGTGAGGCCACCCACCGCAAGT ACTATTCCACCACCCAGGTGCTGTGCTGTCGGCCAGCGTGAGGCTGCCCGCCTCAT TGAGATGTACCGAGACCTCTTCCAGCAGGGGACCTGA		
	ORF Start: at 2		ORF Stop: TGA at 1601

	SEQ ID NO: 172	533 aa	MW at 59379.2kD
NOV13c, 246864086 Protein Sequence	TMQSCSSGDSADDPLSRGLRRRGQPRVVVIGAGLAGLAAAKALLEQGFTDVTVLEAS SHIGGRVQSVKLGHATFELGATWIHGSHGNPIYHLAEANGLLEETTDGERSVGRISLY SKNGVACYLTNHGRRIPKDVVEEFSDLYNEVYNLTQEFFRHDKPVNAESQNSVGVFTR EEVRNRIRNDPDDPEATKRLKLAMIQQYLKVESCESSSSHSMDEVSLSAFGWEIIPGA HHIIIPSGFMRVVELLAEGI PAHVIQLGKPVRCIHWDAQASARPRGPEIEPRGVLRQYT SFFRPGLPTEKVAAIHRLGIGTTDKIFLEFEEFPWGPCNSLQFVWEDEAESHTLTYP PELWYRKICGFDVLYPPERYGHVLSGWICGEEALVMEKCDDEAVAEICTEMLRQFTGN PNIPKPRRILRSAWGSPYFRGSYSYTVQVSSGADVEKLAKPLPYTESSKTAHGSSTK QQPGHLFSSKCEQPLDANRGAVKPMQVLFSGEATHRKYYSTHALLSQGREARLI EMYRDLFQOGT		
	SEQ ID NO: 173	1693 bp	
NOV13d, 258280083 DNA Sequence	CACCATGGGACATCATCACCACCATCACCAAAGTTGTGAATCCAGTGGTGACAGTGCG GATGACCCCTCTCAGTCGCGGCTACGGAGAAGGGGACAGCCTCGTGTGGTGGTGATCG GCGCCGGCTTGGCTGGCTGGCTGCAGCCAAAGCACTTCTTGAGCAGGGTTTCACGGA TGTCAC'TGTGCTTGAGGCTTCCAGCCACATCGGAGGCCGTGTGCAGAGTGTGAAACTT GGACACGCCACCTTTGAGCTGGGAGCCACCTGGATCCATGGCTCCCATGGGAACCCCTA TCTATCATCTAGCAGAAGCCAACGGCTCCTGGAAGAGACAACCGATGCCAGCGCAG CGTGGGCCGATCAGCCTCTATTCCAAGATGGCGTGGCTGCTACCTTACCAACCAC GGCCGAGGATCCCCAAGGACGTGGTTGAGGAATTGAGCGATTTATACAACGAGGTCT ATAAC'TGACCCAGGAGTCTTCCGGCACGATAAACAGTCAATGCTGAAAGTCAAAA TAGCGTGGGGGTGTTACCCGAGAGGAGGTGCGTAACCGCATCAGGAATGACCCTGAC GACCCAGAGGTACCAAGCGCTGAAGCTCGCATGATCCAGCAGTACCTGAAGGTGG AGAGCTGTGAGAGCAGCTCACACAGCATGGACGAGGTGTCCCTGAGCGCTTCGGGGA GTGGACCGAGATCCCCGGCGCTCACACATCATCCCTCGGGCTTCATGCGGGTTGTG GAGCTGCTGGCGGAGGGCATCCCTGCCCACGTATCCAGTAGGGAAACCTGTCCGCT GCATTCAGTGGGACAGGCCTCAGCCCGCCCCAGAGGCCCTGAGATTGAGCCCCGGGG TGAGGGCGACCAATCACGACACTGGGGAGGTTGGCCAGGGTGGAGAGGAGCCCCGG GGGGGCAGGTGGGATGAGGATGAGCAGTGGTGGTGGTGGTGGTGGTGGTGGTGGT AGCTGATCCCGCGGACCATGTGATTGTGACCGTGTGCTAGGTGTGCTAAAGAGGCA GTACACAGTTTCTTCCGGCCAGGCCTGCCACAGAGAAGGTGGCTGCCATCCACCGC CTGGGCATTGGCACCACCGACAAGATCTTTCTGGAATTCGAGGAGCCCTTCTGGGGCC CTGAGTGCAACAGCCTACAGTTTGTGTGGGAGGACGAAGCAGAGGCCACACCCCTCAC CTACCCACCTGAGCTCTGGTACCGCAAGATCTGCGGCTTGTATGCTCTCTACCCGCT GAGCGCTACGGCCATGTGCTGAGCGGCTGGATCTGCGGGGAGGAGGCCCTCGTCATGG AGAAGTGTGATGACGAGGAGTGGCCGAGATCTGCACGGAGATGCTGCGTCAGTTCAC AGGGAACCCCAACATTCCAAAACCTCGGCGAATCTTGCCTCGGCCCTGGGGCAGCAAC CCTTACTTCCGCGGCTCCTATTATACACGAGGTGGGCTCCAGCGGGGCGGATGTGG AGAAGCTGGCCAAGCCCTGCCGTACACGAGAGCTCAAAGACAGGCCCATGTCAGGT GCTGTTTCCGGTGGAGGCCACCCACGCAAGTACTATTCCACCACCCACGGTGTCTG CTGTCCGGCCAGCGTGGGCTGCCCGCTCATTGAGATGTACCGAGACCTCTTCCAGC AGGGGACCTGA		
	ORF Start: at 2		ORF Stop: TGA at 1691
	SEQ ID NO: 174	563 aa	MW at 62799.6kD
NOV13d, 258280083 Protein Sequence	TMGHHHHHHQSCSSGDSADDPLSRGLRRRGQPRVVVIGAGLAGLAAAKALLEQGFTD VTVLEASSHIGGRVQSVKLGHATFELGATWIHGSHGNPIYHLAEANGLLEETTDGERS VGRISLYSKNGVACYLTNHGRRIPKDVVEEFSDLYNEVYNLTQEFFRHDKPVNAESQN SVGVFTREEVRNRIRNDPDDPEATKRLKLAMIQQYLKVESCESSSSHSMDEVSLSAFGE WTEIPGAHHIIIPSGFMRVVELLAEGI PAHVIQLGKPVRCIHWDAQASARPRGPEIEPRG EGDHNDTGEQGQGGEEPRGGRWDEDEQWSVVVECEDCELI PADHVI VTVSLGLVKRQ YTSFFRPGLPTEKVAAIHRLGIGTTDKIFLEFEEFPWGPCNSLQFVWEDEAESHTLT YPPELWYRKICGFDVLYPPERYGHVLSGWICGEEALVMEKCDDEAVAEICTEMLRQFT GNPNI PKPRRILRSAWGSPYFRGSYSYTVQVSSGADVEKLAKPLPYTESSKTA PMQV LFSGEATHRKYYSTHALLSQGREARLI EMYRDLFQOGT		
	SEQ ID NO: 175	1672 bp	
NOV13e, 258280086 DNA	CACCATGCAAAGTTGTGAATCCAGTGGTGACAGTGGGATGACCCCTCTCAGTCGCGGC CTACGGAGAAGGGACAGCCTCGTGTGGTGGTGTGATCGGCGCCGGCTTGGCTGGCCTGG		

258280066 DNA Sequence	CTGCAGCCAAAGCACTTCTTGAGCAGGGTTTCACGGATGTCACTGTGCTTGAGGCTTC CAGCCACATCGGAGGCCGTGTGCAGAGTGTGAAACTTGGACACGCCACCTTTGAGCTG GGAGCCACCTGGATCCATGGCTCCCATGGGAACCTATCTATCATCTAGCAGAAGCCA ACGGCCTCCTGGAAGAGACAACCGATGGGGAACGCAGCGTGGGCCGATCAGCCTCTA TTCCAAGAATGGCGTGGCCTGTACCTTACCAACCACGGCCGAGGATCCCCAAGGAC GTGGTTGAGGAATTCAGCGATTATACAACGAGGTCTATAACTTGACCCAGGAGTTCT TCCGGCACGATAAACCACTCAATGCTGAAAGTCAAAATAGCGTGGGGGTGTTACCCCG AGAGGAGGTGCGTAACCGCATCAGGAATGACCCTGACGACCCAGAGGCTACCAAGCGC CTGAAGCTCGCCATGATCCAGCAGTACCTGAAGGTGGAGAGCTGTGAGAGCAGCTCAC ACAGCATGGACGAGGTGTCCCTGAGCGCCTTCGGGGAGTGGACCGAGATCCCCGGCGC TCACCACATCATCCCTCGGGCTTCATGCGGGTTGTGGAGCTGCTGGCGGAGGGCATC CCTGCCACGTCATCCAGCTAGGGAAACCTGTCCGCTGCATTCACTGGGACAGGCCT CAGCCCGCCCCAGAGGCCCTGAGATTGAGCCCCGGGGTGGAGGCGACCACAATCACGA CACTGGGGAGGGTGGCCAGGGTGGAGAGGAGCCCCGGGGGGGAGGTGGGATGAGGAT GAGCAGTGGTTCGGTGGTGGTGGAGTGCAGGACTGTGAGCTGATCCCGGCGGACCATG TGATTGTGACCGTGTGCTAGTGTGCTAAAGAGGCAGTACACCAAGTTTCTTCCGGCC AGGCCTGCCCCACAGAGAAGGTGGCTGCCATCCACCGCTGGGCATTGGCACCACCGAC AAGATCTTTCTGGAATTCGAGGAGCCCTTCTGGGGCCCTGAGTGCACACAGCTACAGT TTGTGTGGGAGGACGAAGCAGAGAGCCACACCTCACCTACCCACCTGAGCTCTGGTA CCGCAAGATCTGCGGCTTTGATGTCCTCTACCCGCTGAGCGCTACGCCCATGTGCTG AGCGGCTGGATCTGCGGGGAGGAGGCCCTCGTCATGGAGAAGTGTGATGACGAGGCAG TGGCCGAGATCTGCACGGAGATGTGCGTCAGTTTACAGGGAACCCCAACATTCACAAA ACCTCGGCGAATCTGCGCTCGGCCTGGGGCAGCAACCTTACTTCCGCGGCTCCCTAT TCATACACGCAGGTGGGCTCCAGCGGGGCGGATGTGGAGAAGCTGGCCAAGCCCCGTC CGTACACGGAGAGCTCAAAGACAGCGCCCATGCAGGTGCTGTTTTCCGGTGAGGCCAC CCACCGCAAGTACTATTCCACCACCCACGGTGTCTGTGTCCGGCCAGCGTGAGGCT GCCCGCTCATTGAGATGTACCGAGACCTCTTCCAGCAGGGGACCTGA		
	ORF Start: at 2		ORF Stop: TGA at 1670
	SEQ ID NO: 176	556 aa	MW at 61919.7kD
NOV13e, 258280066 Protein Sequence	TMQSCSSGDSADDPRLRRGQPRVVVIGAGLAGLAAAKALLEQGFDTVTVLEAS SHIGGRVQSVKLGHATFELGATWIHGSHGNPIYHLAEANGLLEETDGRSVGRISLY SKNGVACYLTNHGRRIPKDVVEEFSPLYNEVYNLTQEFFRHDKPVNAESQNSVGVFTR EEVRNRIRNDPDDPEATKRLKLAMIQQYLKVSESSSHSMDEVSLSAFGEWTEIPGA HHIIIPSGFMRVVELLAEGI PAHVIQLGKPVRCIHWDQASARPRGPEIEPRGEGDHNHD TEGGGQGGEEPGRGRWDEDEQWSVVVECEDCELI PADHVI VTVSLGLVKRQYTSFFRP GLPTEKVAATHRLGIGTTDKI FLEFEEFPWGPECNSLQFVWEDEABSHLTYPPELWY RKICGFDVLYPPERYGHVLSGWCGEALVMEKCDDEA VAEICTEMLRQFTGNPNIPK PRRILRSAWSNPYFRGSYSYTVQVSSGADVEKLAKPLPYTESSKTAPMQVLFSGEAT HRKYYSTHGAALLSGOREAARLIEMYRDLFQOGT		
	SEQ ID NO: 177	1690 bp	
NOV13f, 258329988 DNA Sequence	CACCATGCAAAGTTGTGAATCCAGTGGTGACAGTGCAGGATGACCTCTCAGTCGCGGC CTACGGAGAAGGGGACAGCCTCGTGTGGTGGTGATCGGCGCCGGCTTGGCTGGCCTGG CTGCAGCCAAAGCACTTCTTGAGCAGGGTTTCACGGATGTCACTGTGCTTGAGGCTTC CAGCCACATCGGAGGCCGTGTGCAGAGTGTGAAACTTGGACACGCCACCTTTGAGCTG GGAGCCACCTGGATCCATGGCTCCCATGGGAACCTATCTATCATCTAGCAGAAGCCA ACGGCCTCCTGGAAGAGACAACCGATGGGGAACGCAGCGTGGGCCGATCAGCCTCTA TTCCAAGAATGGCGTGGCCTGTACCTTACCAACCACGGCCGAGGATCCCCAAGGAC GTGGTTGAGGAATTCAGCGATTATACAACGAGGTCTATAACTTGACCCAGGAGTTCT TCCGGCACGATAAACCACTCAATGCTGAAAGTCAAAATAGCGTGGGGGTGTTACCCCG AGAGGAGGTGCGTAACCGCATCAGGAATGACCCTGACGACCCAGAGGCTACCAAGCGC CTGAAGCTCGCCATGATCCAGCAGTACCTGAAGGTGGAGAGCTGTGAGAGCAGCTCAC ACAGCATGGACGAGGTGTCCCTGAGCGCCTTCGGGGAGTGGACCGAGATCCCCGGCGC TCACCACATCATCCCTCGGGCTTCATGCGGGTTGTGGAGCTGCTGGCGGAGGGCATC CCTGCCACGTCATCCAGCTAGGGAAACCTGTCCGCTGCATTCACTGGGACAGGCCT CAGCCCGCCCCAGAGGCCCTGAGATTGAGCCCCGGGGTGGAGGCGACCACAATCACGA CACTGGGGAGGGTGGCCAGGGTGGAGAGGAGCCCCGGGGGGGAGGTGGGATGAGGAT GAGCAGTGGTTCGGTGGTGGTGGAGTGCAGGACTGTGAGCTGATCCCGGCGGACCATG TGATTGTGACCGTGTGCTAGTGTGCTAAAGAGGCAGTACACCAAGTTTCTTCCGGCC		

	AGGCCTGCCACAGAGAAGGTGGCTGCCATCCACCGCCTGGGCATTGGCACCACCGAC AAGATCTTTCTGGAATTCGAGGAGCCCTTCTGGGGCCCTGAGTGCAACAGCCTACAGT TTGTGTGGGAGGACGAAGCAGAGAGCCACACCCTCACCTACCCACCTGAGCTCTGGTA CCGCAAGATCTGCGGCTTTGATGTCTCTACCCGCTGAGCGCTACGGCCATGTGCTG AGCGGCTGGATCTGCGGGGAGGAGGCCCTCGTCATGGAGAAGTGTGATGACGAGGCAG TGGCCGAGATCTGCACGAGATGCTGCGTCAGTTCACAGGGAACCCCAACATCCAAA ACCTCGGCGAATCTTGCCTCGGCCCTGGGCAGCAACCCTTACTTCCGCGGCTCTAT TCATACACGAGGTGGGCTCCAGCGGGGCGGATGTGGAGAAGCTGGCCAAGCCCTGC CGTACACGAGAGCTCAAAGACAGCGCCCATGCAGGTGCTGTTTTCCGGTGAGGCCAC CCACCGCAAGTACTATTCCACCACCCACGGTGTCTGTGCTGTCGGCCAGCGTGAGGCT GCCCGCTCATTGAGATGTACCGAGACCTCTTCCAGCAGGGGACCCATCATACCACC ATCACTGA		
	ORF Start: at 2		ORF Stop: TGA at 1688
	SEQ ID NO: 178	562 aa	MW at 62742.6kD
NOV13f, 258329988 Protein Sequence	TMQSCSSGDSADDPLSRGLRRRGQPRVVVIGAGLAGLAAAKALLEQGFTDVTVLEAS SHIGGRVQSVKLGHATFELGATWIGHSHGNPIYHLAEANGLLEETDGDERSVGRISLY SKNGVACYLTNHGRRIPKDVVEEFSLDYNEVYNLTQEFFRHDKPVNAESQNSVGVFTR BEVRNRIRNDPDDPEATKRLKLANIQQYLKVESCESSSHSMDEVSLSAFGEWTEIPGA HHIIPSGFMRVVELLAEGI PAHVILQKPVRCIHWDAQASARPRGPEIEPRGEGDHNHD TGEGGGQGEEPRGGRWDEDEQWSVVVECEDCELI PADHVIVTVSLGVLKRQYTSFFRP GLPTEKVAAIHRLGIGTTDKIFLEFEEPPFWGPECNSLQFVWEDEAESHTLTYPPELWY RKICGFDVLYPPERYGHVLSGWICGEEALVMEKCDDEAVAIECTEMLRQFTGNPNIPK PRRILRSAWGSPYFRGSYSYTVQVSSGADVEKLAKPLPYTESKTA PMQVLFSGEAT HRKYYSTTHGALLSGQREARLIEMYRDLFQQGTHHHHHH		
	SEQ ID NO: 179	1700 bp	
NOV13g, 254047897 DNA Sequence	AAGGAAAAAAGCGGCCGCCACCATGCAAAGTTGTGAATCCAGTGGTGACAGTGCGGAT GACCCCTCTCAGTCGCGGCCTACGGAGAAGGGGACAGCCTCGTGTGGTGGTGATCGGCG CCGGCTTGGCTGGCCTGGCTGCAGCCAAAGCACTTCTTGAGCAGGGTTTACGGATGT CACTGTGCTTGAGGCTTCCAGCCACATCGGAGGCCGTGTGCAGAGTGTGAACTTGGA CAGCCACCTTTGAGCTGGGAGCCACCTGGATCCATGGCTCCCATGGGAACCCATATCT ATCATCTAGCAGAAGCCAACGGCTCTCTGGAAGAGACAACCGATGGGGAACGCAGCGT GGGCCGCATCAGCCTCTATTCCAAGAATGGCGTGGCCTGCTACCTTACCAACCACGGC CGCAGGATCCCCAAGGACGTGGTTGAGGAATTCAGCGATTTATACAACGAGGTCTATA ACTTGACCCAGGAGTTCTTCCGGCAGGATAAACCAGTCAATGCTGAAAGTCAAATAG CGTGGGGGTGTTACCCGAGAGGAGGTGCGTAACCGCATCAGGAATGACCTGACGAC CCAGAGGTACCAAGCGCTGAAGCTCGCCATGATCCAGCAGTACCTGAAGGTGGAGA GCTGTGAGACGAGCTCACACAGCATGGACGAGGTGTCCTGAGCGCCTTCGGGAGTG GACCGAGATCCCCGGCGCTCACCACATCATCCCTCGGGCTTCATGCGGGTTGTGGAG CTGCTGGCGGAGGGCATCCCTGCCACGTATCCAGCTAGGGAAACCTGTCCGCTGCA TTCACTGGGACCAGGCCTCAGCCCGCCCCAGAGCCCTGAGATTGAGCCCCGGGGTGA GGGCGACCACAATCACGACACTGGGGAGGGTGGCCAGGGTGGAGAGGAGCCCCGGGGG GGCAGGTGGGATGAGGATGAGCAGTGGTTCGGTGGTGGTGGAGTGCGAGGACTGTGAGC TGATCCCGCGGACCATGTGATTGTGACCGTGTGCTAGGTGTGCTAAAGAGGCAGTA CACCAGTTTCTTCCGGCCAGGCCTGCCACAGAGAAGGTGGCTGCCATCCACGCCTG GGCATTGGCACCACCGACAAGATCTTTCTGGAATTCGAGGAGCCCTTCTGGGGCCCTG AGTGCAACAGCCTACAGTTTGTGTGGGAGGACGAAGCAGAGAGCCACACCTCACCTA CCCACCTGAGCTCTGGTACCGCAAGATCTGCGGCTTTGATGTCCTCTACCCGCTGAG CGCTACGGCCATGTGCTGAGCGGCTGGATCTGCGGGGAGGAGGCCCTCGTCATGGAGA AGTGTGATGACGAGGCAGTGGCCGAGATCTGCACGGAGATGCTGCGTCAGTTCACAGG GAACCCCAACATTCCAAACCTCGGCGAATCTTGCCTCGGCTGGGGCAGCAACCCT TACTTCCGCGGCTCTATTATACACGAGGTGGGCTCCAGCGGGGCGGATGTGGAGA AGCTGGCCAAGCCCTGCCGTACACGGAGAGCTCAAAGACAGCGCCCATGCAGGTGCT GTTTTCCGGTGAGGCCACCCACCGCAAGTACTATCCACCACCCACCGTGTCTGTGCTG TCCGGCCAGCGTGAGGCTGCCCGCTCATTGAGATGTACCGAGACCTCTTCCAGCAGG GGACCTGATCTAGACTAG		
	ORF Start: at 2		ORF Stop: TGA at 1688
	SEQ ID NO: 180	562 aa	MW at 62545.5kD

NOV13g, 254047897 Protein Sequence	RKKAAATMQSCSSGDSADDP LSRGLRRRGQPRVVVIGAGLAGLAAAKALLEQGFTDV TVLEASSHIGGRVQSVKLGHATFELGATWIHGSHGNPIYHLAEANGLLEETTDGERSV GRISLYSKNGVACYLTNHGRRIPKDVVEEFSDLYNEVYNLTQEFFRHDKPVNAESQNS VGVFTR E EVNRIRNDPDDPEATKRLKLAMIQQYLKVESCESSSHSMDEVSLSAFGEW TEIPGAHHIIPSGFMRVVELLAEGIPAHVIQLGKPVRCIHWDQASARPRGPEIEPRGE GDHNHDTGEGGQGGEEPRGGRWDEDEQWSVVVECEDCELI PADHVIVTVSLGLVKRQY TSFFRPGLPTEKVAAIHRLLGIGTTDKIFLEFEEFPWGPENSLQFVWEDEAESHTLT PPELWYRKICGFDVLYPPERYPYGHVLSGWICGEEALVMEKCDDEAVAEICTEMLRQFTG NPNI PKPRRILRSAWGSPYFRGSYSYTVQVSSGADVEKLAKPLPYTESSKTAPMQVL FSGEATHRKYYSTTHGALLSGQREAAARLIEMYRDLFQQGT		
	SEQ ID NO: 181	1690 bp	
NOV13h, 258329988 DNA Sequence	CACCATGCAAAGTTGTGAATCCAGTGGTGACAGTGCGGATGACCCTCTCAGTCGCGGC CTACGAGAAGGGGACAGCCTCGTGTGGTGGTGATCGGCGCCGGCTTGGCTGGCCTGG CTGCAGCCAAAGCACTTCTTGAGCAGGGTTTACGGATGTCACTGTGCTTGAGGCTTC CAGCCACATCGGAGGCGGTGTGCAGAGTGTGAACTTGACACGCCACCTTTGAGCTG GGAGCCACCTGGATCCATGGCTCCCATGGGAACCCATATCATCTAGCAGAAGCCA ACGGCTCTCTGGAAGAGACAACCGATGGGGAACGCAGCGTGGGCGCATCAGCCTCTA TTCCAAGAATGGCGTGGCCTGCTACCTTACCAACCACGCGCGCAGGATCCCCAAGGAC GTGGTTGAGGAATTCAGCGATTATACAACGAGGTCTATAACTTGACCCAGGAGTTCT TCCGGCAGGATAAACCAGTCAATGCTGAAAGTCAAATAGCGTGAGGGGTGTTACCCG AGAGAGGTGCGTAACCGCATCAGGAATGACCCTGACGACCCAGGCTACCAAGCGC CTGAAGCTCGCCATGATCCAGCAGTACCTGAAGGTGGAGAGCTGTGAGAGCAGCTCAC ACAGCATGGACGAGGTGTCCCTGAGCGCCTTCGGGAGTGGACCGAGATCCCCGGCGC TCACCACATCATCCCTCGGGCTTCATGCGGGTTGTGGAGCTGTGGCGGAGGGCATC CCTGCCACGTCATCCAGCTAGGGAACCTGTCCGCTGCATTCACTGGGACAGGCCT CAGCCCGCCCCAGAGGCCCTGAGATTGAGCCCGGGGTGAGGGCGACCAACAATCAGCA CACTGGGAGGGTGGCCAGGGTGGAGAGGAGCCCCGGGGGGCAGGTGGGATGAGGAT GAGCAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG TGATTGTGACCGTGTGCTAGGTGTGCTAAAGAGGCAGTACACAGTTTCTTCCGGCC AGGCCTGCCCACAGAGAAGGTGGCTGCCATCCACCGCCTGGGCATTGGCACCACCGAC AAGATCTTTCTGGAATTCGAGGAGCCCTTCTGGGGCCCTGAGTGCAACAGCCTACAGT TTGTGTGGGAGGACGAAGCAGAGAGCCACACCTCACCTACCCACCTGAGCTCTGGTA CCGCAAGATCTGCGGCTTTGATGTCTCTACCCGCTGAGCGCTACGGCCATGTGCTG AGCGGTGGATCTGCGGGAGGAGGCCCTCGTCATGGAGAAGTGTGATGACGAGGCAG TGGCCGAGATCTGCACGGAGATGCTGCGTCAGTTACAGGGAACCCCAACATTCACAAA ACCTCGGCAATCTTGGCTCGGCCCTGGGGCAGCAACCTTACTTCCGGGCTCCTAT TCATACACGAGGTGGGCTCCAGCGGGCGGATGTGGAGAAGCTGGCCAAGCCCCTGC CGTACACGGAGAGCTCAAAGACAGCGCCCATGCAGGTGCTGTTTTCCGGTGAGGCCAC CCACCGCAAGTACTATTCCACCACCCACGGTGTCTGTCTGCTGTCGGCCAGCGTGAAGCT GCCCGCTCATTTAGATGTACCAGACCTCTTCCAGCAGGGGACCATCATCACCACC ATCACTGA		
	ORF Start: at 2		ORF Stop: TGA at 1688
	SEQ ID NO: 182	562 aa	MW at 62742.6kD
NOV13h, 258329988 Protein Sequence	TMQSCSSGDSADDP LSRGLRRRGQPRVVVIGAGLAGLAAAKALLEQGFTDVTVLEAS SHIGGRVQSVKLGHATFELGATWIHGSHGNPIYHLAEANGLLEETTDGERSVGRISLY SKNGVACYLTNHGRRIPKDVVEEFSDLYNEVYNLTQEFFRHDKPVNAESQNSVGVFTR EEVNRIRNDPDDPEATKRLKLAMIQQYLKVESCESSSHSMDEVSLSAFGEWTEIPGA HHIIPSGFMRVVELLAEGIPAHVIQLGKPVRCIHWDQASARPRGPEIEPRGEGDHNH TGEGGQGGEEPRGGRWDEDEQWSVVVECEDCELI PADHVIVTVSLGLVKRQYTSFFRP GLPTEKVAAIHRLLGIGTTDKIFLEFEEFPWGPENSLQFVWEDEAESHTLTYPPELWY RKICGFDVLYPPERYPYGHVLSGWICGEEALVMEKCDDEAVAEICTEMLRQFTGNPNI PK PRRILRSAWGSPYFRGSYSYTVQVSSGADVEKLAKPLPYTESSKTAPMQVLFSGEAT HRKYYSTTHGALLSGQREAAARLIEMYRDLFQQGTHHHHHH		
	SEQ ID NO: 183	1672 bp	
NOV13i, 258280066 DNA Sequence	CACCATGCAAAGTTGTGAATCCAGTGGTGACAGTGCGGATGACCCTCTCAGTCGCGGC CTACGAGAAGGGGACAGCCTCGTGTGGTGGTGATCGGCGCCGGCTTGGCTGGCCTGG CTGCAGCCAAAGCACTTCTTGAGCAGGGTTTACGGATGTCACTGTGCTTGAGGCTTC CAGCCACATCGGAGGCGGTGTGCAGAGTGTGAACTTGACACGCCACCTTTGAGCTG GGAGCCACCTGGATCCATGGCTCCCATGGGAACCCATATCATCTAGCAGAAGCCA ACGGCTCTCTGGAAGAGACAACCGATGGGGAACGCAGCGTGGGCGCATCAGCCTCTA TTCCAAGAATGGCGTGGCCTGCTACCTTACCAACCACGCGCGCAGGATCCCCAAGGAC GTGGTTGAGGAATTCAGCGATTATACAACGAGGTCTATAACTTGACCCAGGAGTTCT TCCGGCAGGATAAACCAGTCAATGCTGAAAGTCAAATAGCGTGAGGGGTGTTACCCG AGAGAGGTGCGTAACCGCATCAGGAATGACCCTGACGACCCAGGCTACCAAGCGC CTGAAGCTCGCCATGATCCAGCAGTACCTGAAGGTGGAGAGCTGTGAGAGCAGCTCAC ACAGCATGGACGAGGTGTCCCTGAGCGCCTTCGGGAGTGGACCGAGATCCCCGGCGC TCACCACATCATCCCTCGGGCTTCATGCGGGTTGTGGAGCTGTGGCGGAGGGCATC CCTGCCACGTCATCCAGCTAGGGAACCTGTCCGCTGCATTCACTGGGACAGGCCT CAGCCCGCCCCAGAGGCCCTGAGATTGAGCCCGGGGTGAGGGCGACCAACAATCAGCA CACTGGGAGGGTGGCCAGGGTGGAGAGGAGCCCCGGGGGGCAGGTGGGATGAGGAT GAGCAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG TGATTGTGACCGTGTGCTAGGTGTGCTAAAGAGGCAGTACACAGTTTCTTCCGGCC AGGCCTGCCCACAGAGAAGGTGGCTGCCATCCACCGCCTGGGCATTGGCACCACCGAC AAGATCTTTCTGGAATTCGAGGAGCCCTTCTGGGGCCCTGAGTGCAACAGCCTACAGT TTGTGTGGGAGGACGAAGCAGAGAGCCACACCTCACCTACCCACCTGAGCTCTGGTA CCGCAAGATCTGCGGCTTTGATGTCTCTACCCGCTGAGCGCTACGGCCATGTGCTG AGCGGTGGATCTGCGGGAGGAGGCCCTCGTCATGGAGAAGTGTGATGACGAGGCAG TGGCCGAGATCTGCACGGAGATGCTGCGTCAGTTACAGGGAACCCCAACATTCACAAA ACCTCGGCAATCTTGGCTCGGCCCTGGGGCAGCAACCTTACTTCCGGGCTCCTAT TCATACACGAGGTGGGCTCCAGCGGGCGGATGTGGAGAAGCTGGCCAAGCCCCTGC CGTACACGGAGAGCTCAAAGACAGCGCCCATGCAGGTGCTGTTTTCCGGTGAGGCCAC CCACCGCAAGTACTATTCCACCACCCACGGTGTCTGTCTGCTGTCGGCCAGCGTGAAGCT GCCCGCTCATTTAGATGTACCAGACCTCTTCCAGCAGGGGACCATCATCACCACC ATCACTGA		

	CAGCCACATCGGAGGCCGTGTGCAGAGTGTGAACTTGGACACGCCACCTTTGAGCTG GGAGCCACCTGGATCCATGGCTCCCATGGGAACCTATCTATCATCTAGCAGAAGCCA ACGGCCTCCTGGAAGAGACAACCGATGGGGAACGCAGCGTGGGCCGCATCAGCCTCTA TTCCAAGAATGGCGTGGCCTGCTACCTTACCAACCACGGCCGCAGGATCCCCAAGGAC GTGGTTGAGGAATTCAGCGATTTATACAACGAGGTCTATAACTTGACCCAGGAGTTCT TCCGGCACGATAAACCAGTCAATGCTGAAAGTCAAAATAGCGTGGGGGTGTTCACCCG AGAGGAGGTGCGTAACCGCATCAGGAATGACCTTGACGACCCAGAGGCTACCAAGCGC CTGAAGCTCGCCATGATCCAGCAGTACCTGAAGGTGGAGAGCTGTGAGAGCAGCTCAC ACAGCATGGACGAGGTGTCCCTGAGCGCCTTCGGGGAGTGGACCGAGATCCCCGGCGC TCACCACATCATCCCCTCGGGCTTCATGCGGGTTGTGGAGCTGTGTCGGGAGGGGCATC CCTGCCACAGTCATCCAGCTAGGGAAACCTGTCCGCTGCATTCACTGGGACCAGGCCCT CAGCCCGCCCCAGAGGCCCTGAGATTGAGCCCCGGGGTGAGGGCGACCAACAATCACGA CACTGGGGAGGGTGGCCAGGGTGGAGAGGAGCCCCGGGGGGCAGGTGGGATGAGGAT GAGCAGTGGTCCGTGGTGGTGGAGTGGGAGGACTGTGAGCTGATCCCGCGGACCATG TGATTGTGACCGTGTGCTAGGTGTGCTAAAGAGGCAGTACACCAAGTTCTTCCGGCC AGGCCTGCCACAGAGAAGGTGGCTGCCATCCACCGCCTGGGCATTGGCACCACCGAC AAGATCTTTCTGGAATTCGAGGAGCCCTTCTGGGGCCCTGAGTGCAACAGCCTACAGT TTGTGTGGGAGGACGAAGCAGAGAGCCACACCTCACCTACCCACCTGAGCTCTGGTA CCGCAAGATCTGCGGCTTTGATGTCTCTACCCGCCTGAGCGCTACGGCCATGTGCTG AGCGGCTGGATCTGCGGGGAGGAGGCCCTCGTCATGGAGAAGTGTGATGACGAGGCAG TGGCCGAGATCTGCACGGAGATGTGCGTCAGTTCACAGGGAACCCCAACATTCCAAA ACCTCGGCGAATCTTCCGCTCGGCCTGGGGCAGCAACCCTTACTTCCGGCGCTCCTAT TCATACACGCAAGGTGGCTCCAGCGGGCGGATGTGGAGAAGCTGGCCAAGCCCCCTGC CGTACACGGAGAGCTCAAAGACAGCGCCCATGCAGGTGCTGTTTTCCGGTGAGGCCAC CCACCGCAAGTACTATTCACCACCCACGGTCTCTGCTGTCCGGCCAGCGTGAGGCT GCCCCCTCATTGAGATGTACCGAGACCTCTTCCAGCAGGGGACCTGA		
	ORF Start: at 2		ORF Stop: TGA at 1670
	SEQ ID NO: 184	556 aa	MW at 61919.7kD
NOV13i, 258280066 Protein Sequence	TMQSCSSGDSADDPLSRGLRRRQPRVVVIGAGLAGLAAAKALLEQFTDVTVLEAS SHIGGRVQSVKLGHATFELGATWIHGSHGNPIYHLAEANGLLEETDGRSVGRISLY SKNGVACYLTNHGRRIPKDVVEEFSDLNEVYNLTQEFFRHDKPVNAESQNSVGVFTR EEVRNRIRNDPDDPEATKRLKLAMIQQYLKVSESSSSHSMDEVSLSAFGEWTEIPGA HHIIIPSGFMRVVELLAEGIPAHVILQKGPVRCIHWDAQASARPRGPEIEPRGEGDHNHD TGGGGQGGEEPRGGRWDEDEQWSVVVECEDCELI PADHVI VTVSLGLVKRQYTSFFRP GLPTEKVAAIHRLGIGTTDKIFLEFEEPPFWGPECNSLQFVWEDEAESHTLTYPPELWY RKICGFDVLYPPERYGHVLSGWI CGEEALVMEKCDDEAVAEICTEMLRQFTGNPNIPK PRRIILRSAWSNPYFRGSYSYTVQVSSGADVEKLAKPLPYTESSKTPMQVLFSGEAT HRKYYSTTHGALLSGQREARLIEMYRDLFQQGT		
	SEQ ID NO: 185	1693 bp	
NOV13j, 258280083 DNA Sequence	CACCATGGGACATCATCACCACCATCACCAGTTGTGAATCCAGTGGTGACAGTGGC GATGACCTCTCAGTCGCGGCCTACGAGAAGGGGACAGCCTCGTGTGGTGGTGATCG GCGCCGCTTGGCTGGCCTGGCTGCAGCCAAAGCACTTCTTGAGCAGGGTTTCACGGA TGTCACTGTGCTTGAGGCTTCCAGCCACATCGGAGGCCGTGTGCAGAGTGTGAACTT GGACACGCCACCTTTGAGCTGGGAGCCACCTGGATCCATGGCTCCCATGGGAACCTTA TCTATCATCTAGCAGAAGCCAACGGCCTCCTGGAAGAGACAACCGATGGGGAACGCAG CGTGGGCCGCATCAGCCTCTATTCCAAGAATGGCGTGGCCTGCTACCTTACCAACCAC GGCCGCAGGATCCCCAAGGACGTGGTTGAGGAATTCAGCGATTTATACAACGAGGTCT ATAACTTGACCCAGGAGTTCTTCCGGCACGATAAACCAGTCAATGCTGAAAGTCAAAA TAGCGTGGGGGTGTTACCCGAGAGGAGGTGCGTAACCGCATCAGGAATGACCTGAC GACCCAGAGGCTACCAAGCGCCTGAAGCTCGCCATGATCCAGCAGTACCTGAAGGTGG AGAGCTGTGAGAGCAGCTCACACAGCATGGACGAGGTGTCCCTGAGCGCCTTCGGGGA GTGGACCGAGATCCCCGGCGCTCACCACATCATCCCCTCGGGCTTCATGCGGGTTGTG GAGCTGCTGGCGGAGGGCATCCCTGCCACGTCATCCAGCTAGGGAAACCTGTCCGCT GCATTCACTGGGACACAGGCCTCAGCCCGCCAGAGGCCCTGAGATTGAGCCCCGGGG TGAGGGCGACCAATCACGACACTGGGGAGGGTGGCCAGGGTGGAGAGGAGCCCCGG GGGGCAGGTGGGATGAGGATGAGCAGTGGTCCGTGGTGGTGGAGTGGCAGGACTGTG AGCTGATCCCGGCGGACCATGTGATTGTGACCGTGTGCTAGGTGTGCTAAAGAGGCA GTACACCAGTTTCTTCCGGCCAGGCCTGCCACAGAGAAGGTGGCTGCCATCCACCGC		

	CTGGGCATTGGCACCACCGACAAGATCTTTCTGGAATTCGAGGAGCCCTTCTGGGGCC CTGAGTGCAACAGCCTACAGTTTGTGTGGGAGGACGAAGCAGAGAGCCACACCCTCAC CTACCCACCTGAGCTCTGGTACCGCAAGATCTGCGGCTTTGATGTCTCTACCCGCCT GAGCGCTACGGCCATGTGCTGAGCGGCTGGATCTGCGGGGAGGAGGCCCTCGTCATGG AGAAGTGTGATGACGAGGCAGTGGCCGAGATCTGCACGGAGATGTGCGTCAGTTCAC AGGGAACCCCAACATTCCAAAACCTCGGCGAATCTTGCGCTCGGCCCTGGGGCAGCAAC CCTTACTTCCGCGGCTCCTATTATACACGCAGGTGGGCTCCAGCGGGCGGATGTGG AGAAGCTGGCCAAGCCCTGCCGTACACGGAGAGCTCAAAGACAGCGCCCATGCAGGT GCTGTTTTCCGGTGAGGCCACCCACCGCAAGTACTATTCCACCACCCACGGTGCTCTG CTGTCCGGCCAGCGTGAGGCTGCCCGCCTCATTGAGATGTACCGAGACCTCTTCCAGC AGGGGACCTGA		
	ORF Start: at 2		ORF Stop: TGA at 1691
	SEQ ID NO: 186	563 aa	MW at 62799.6kD
NOV13j, 258280083 Protein Sequence	TMGHHHHHHQSCSSGDSADDPLSRGLRRRQPRVVVIGAGLAGLAAAKALLEQGFTD VTVLEASSHIGGRVQSVKLGHATFELGATWIHGSHGNPIYHLAEANGLLEETTDGERS VGRISLYSKNGVACYLTNHGRRIPKDVVEEFSPLYNEVYNLTQEFFRHDKPVNAESQN SVGVFTREEVRNRI RNDPDDPEATKRLKLAMIQQYLKVESCESSSHSMDEVSLSAFGE WTEIPGAHHIIPSGFMRVVVELLAEGI PAHVILQKPVRCIHWDQASARPRGPEIEPRG EGDHNHDTGEGGQGGEEPRGRWDEDEQWSVVVECEDCELI PADHVI VTVSLGVLKRQ YTSFFRPLPTEKVAIHRIGITTDKI FLEFEEFPWGPENC SLQFVWEDEAESHTLT YPPELWYRKICGFDVLYPPERYPYGHVLSGWICGEEALVMEKCDDEAVAEICTEMLRQFT GNPNIPKPRRILRSWAGSNPYFRGSYSYTVQVSSGADVEKLAKPLPYTESSKTA PMQV LFSGEATHRKYYSTTHGALLSGQREAAARLIEMYRDLFQQGT		
	SEQ ID NO: 187	1993 bp	
NOV13k, CG140122-02 DNA Sequence	GGCACGAGGGTCCCGCGCGCGCTGGAGGAGGAAGCCAGCGCGCTGGCGGAGGAGGAG AGACGGAGGAGGCGGAGACCGGAGCGCGCTCGCCGAGACTTACTTCCCGGCTCAG CAGGGAAGGTTCTAGAAAGGTGAGCGCGGACGGTATGCAAAGTTGTGAATCCAGTGG TGACAGTGCAGGATGACCTCTCAGTCGCGGCCTACGGAGAAGGGACAGCCTCGTGTG GTGGTGATCGCGCGCGGCTTGGCTGGCCTGGCTGCAGCCAAAGCACTTCTTGAGCAGG GTTTCACGGATGTCACTGTGCTTGAGGCTTCCAGCCACATCGGAGGCCGTGTGCAGAG TGTGAACTTGGACACGCCACCTTTGAGCTGGGAGCCACCTGGATCCATGGCTCCCAT GGGAACCTATCTATCATCTAGCAGAAGCCAACGGCCTCTGGAAGAGACAACCGATG GGGAACGCAGCGTGGGCCGATCAGCCTCTATTCCAAGAAATGGCGTGGCCTGCTACCT TACCAACCACGGCCGAGGATCCCAAGGACGTGGTTGAGGAATTCAGCGATTTATAC AACGAGGTCTATAACTTGACCCAGGAGTTCTTCCGGCACGATAAACAGTCAATGCTG AAAGTCAAAATAGCGTGGGGGTGTTACCCGAGAGGAGGTGCGTAACCGCATCAGGAA TGACCTGACGACCCAGAGGCTACCAAGCGCTGAAGCTCGCCATGATCCAGCATAC CTGAAGGTGGAGAGCTGTGAGAGCAGCTCACACAGCATGGACGAGGTGTCCCTGAGCG CCTTCGGGAGTGGACCGAGATCCCGCGGCTCACCACATCATCCCTCGGGCTTCAT GCGGGTGTGAGAGCTGCTGGCGGAGGGCATCCCTGCCACGTCATCCAGCTAGGGAAA CCTGTCCGCTGCATTCACTGGGACAGGCCTCAGCCGCCCCAGAGGCCCTGAGATTG AGCCCCGGGTGTGCTAAAGAGGCAGTACACCAAGTTTCTTCCGCCAGGCCCTGCCAC AGAGAAGGTGGCTGCCATCCACCGCCTGGGCATTGGCACCACCGACAAGATCTTTCTG GAATTCGAGGAGCCCTTCTGGGGCCCTGAGTGCAACAGCCTACAGTTTGTGTGGGAGG ACGAAGCGGAGAGCCACACCTCACCTACCCACCTGAGCTCTGGTACCGCAAGATCTG CGGCTTTGATGTCTCTACCCGCTGAGCGCTACGGCCATGTGCTGAGCGGCTGGATC TGCGGGGAGGAGGCCCTCGTCATGGAGAAGTGTGATGACGAGGCACTGGCCGAGATCT GCACGGAGATGCTGCGTCACTTCAACAGGAACCCCAACATTCCAAAACCTCGGGGAAT CTTGCGCTCGGCCTGGGGCAGCAACCCTTACTTCCGCGGCTCCTATTATACACGCAG GTGGGCTCCAGCGGGCGGATGTGGAGAAGCTGGCCAAGCCCCCTGCCGTACACGGAGA GCTCAAAGACAGCGCCCATGCAGGTGCTGTTTTCCGGTGAGGCCACCCACCGCAAGTA CTATTCACACCCACGGTGCTGCTGTCCGGCCAGCGTGAGGCTGCCCGCCTCAT GAGATGTACCGAGACCTCTTCCAGCAGGGGACCTGAGGCTGTCTGCTGCTGTGAGAA GAGCCACTAATCGTGACCTCCAGCCTGCCCTTGCTGCCGTGTGCTCGCTTCCCT GATCCTCTGTAGAAAGGATTTTATCTTCTGTAGAGCTAGCCGCCCTGACTGCCTTCA GACCTGGCCCTGTAGCTTTTCTTTTCTCCAGGCTGGGCCGTGAGCAGGTGGGCCGTT GAGTTACCTCTGTGCTGGATCCCGTGCCCCACTTGCTTACCTCTGTCTGCTGCTGT TATTGTAAGTGCTTCAATACTTTGCATTTTGGGATAATAAAAAAGGCTCCCTCCCT		

	GCAAAAAAAAAAAAAAAAAA		
	ORF Start: ATG at 152		ORF Stop: TGA at 1658
	SEQ ID NO: 188	502 aa	MW at 56090.6kD
NOV13k, CG140122-02 Protein Sequence	MQSCSSGDSADDPLSRGLRRRGQPRVVVIGAGLAGLAAAKALLEQGFTDVTVLEASS HIGGRVQSVKLGHATFELGATWIHGSHGNPIYHLAEANGLLEETTDGERSVGRISLYS KNGVACYLTNHGRRIPKDVVEEFSPLYNEVYNLTQEFFRHDKPVNAESQNSVGVFTRE EVRNRIRNDPDDPEATKRLKLAMIQQYLKVESCESSSHSMDEVSLSAFGEWTEIPGAH HIIPSGFMRVVELLAEGIPAHVIQLGKPVRCIHWDAQASARPRGPEIEPRGVLRQYTS FFRPGLPTEKVAAIHRLGIGTTDKIFLEFEEPFWGPENSLQFVWEDEAESHTLTYP ELWYRKICGFDVLYPPERYGHVLSGWICGEEALVMEKCDDEAVAEICTEMLRQFTGNP NIPKPRILRSAGWSNPYFRGSYSYTVQVSSGADVEKLAKPLPYTESSKTAPMQVLF GEATHRKYYSTTHGALLSGQREARLIEMYRDLFQQGT		
	SEQ ID NO: 189	1012 bp	
NOV13l, CG140122-03 DNA Sequence	CACCATGCAAAGTTGTGAATCCAGTGGTGACAGTGCGGATGACCCTCTCAGTCGCGGC CTACGGAGAAGGGGACAGCCTCGTGTGGTGGTGATCGGCGCCGGCTTGCTGGCCTGG CTGCAGCCAAAGCACTTCTTGAGCAGGGTTTCACGGATGTCACTGTGCTTGAGGCTTC CAGCCACATCGGAGGCCGTGTGCAGAGTGTGAACTTGGACACGCCACCTTTGAGCTG GGAGCCACCTGGATCCATGGCTCCCATGGGAACCTATCTATCATCTAGCAGAAGCCA ACGGCCTCCTGGAAGAGACAACCGATGGGGAACGCAGCGTGGGCCGCATCAGCCTCTA TTCCAAGAATGGCGTGGCCTGCTACCTTACCAACCACGGCCGAGGATCCCCAAGGAC GTGGTTGAGGAATTCAGCGATTTATACAACGAGGTCTATACTTGACCCAGGAGTTCT TCCGGCACGATAAACCAGTCAATGCTGAAAGTCAAATAGCGTGGGGGTGTTACCCG AGAGGAGGTGCGTAACCGCATCAGGAATGACCCTGACGACCCAGAGGCTACCAAGCGC CTGAAGCTCGCCATGATCCAGCAGTACCTGAAGGTGGAGAGCTGTGAGAGCAGCTCAC ACAGCATGGACGAGGTGTCCCTGAGCGCCTTCGGGGAGTGGACCGAGATCCCCGGCGC TCACCACATCATCCCTCGGGCTTCATGCGGGTTGTGGAGCTGTGGCGGAGGGCATC CCTGCCACGTCATCCAGCTAGGGAAACCTGTCCGCTGCATTCACTGGGACCAGGCCT CAGCCCGCCCCAGAGGCCCTGAGATTGAGCCCCGCGTACACAGAGAGCTCAAAGAC AGCGCCCATGCAGGTGCTGTTTTCCGGTGAGGCCACCCACGCAAGTACTATTCCACC ACCACGGTGCTCTGCTGTCCGGCCAGCGTGAGGCTGCCCCGCTCATTGAGATGTACC GAGACCTCTTCAGCAGGGGACCTGA		
	ORF Start: at 2		ORF Stop: TGA at 1010
	SEQ ID NO: 190	336 aa	MW at 37093.2kD
NOV13l, CG140122-03 Protein Sequence	TMQSCSSGDSADDPLSRGLRRRGQPRVVVIGAGLAGLAAAKALLEQGFTDVTVLEAS SHIGGRVQSVKLGHATFELGATWIHGSHGNPIYHLAEANGLLEETTDGERSVGRISLYS SKNGVACYLTNHGRRIPKDVVEEFSPLYNEVYNLTQEFFRHDKPVNAESQNSVGVFTRE BEVRNRIRNDPDDPEATKRLKLAMIQQYLKVESCESSSHSMDEVSLSAFGEWTEIPGA HHIIPSGFMRVVELLAEGIPAHVIQLGKPVRCIHWDAQASARPRGPEIEPLPYTESSKT APMQVLFSGEATHRKYYSTTHGALLSGQREARLIEMYRDLFQQGT		
	SEQ ID NO: 191	1603 bp	
NOV13m, CG140122-04 DNA Sequence	CACCATGCAAAGTTGTGAATCCAGTGGTGACAGTGCGGATGACCCTCTCAGTCGCGGC CTACGGAGAAGGGGACAGCCTCGTGTGGTGGTGATCGGCGCCGGCTTGCTGGCCTGG CTGCAGCCAAAGCACTTCTTGAGCAGGGTTTCACGGATGTCACTGTGCTTGAGGCTTC CAGCCACATCGGAGGCCGTGTGCAGAGTGTGAACTTGGACACGCCACCTTTGAGCTG GGAGCCACCTGGATCCATGGCTCCCATGGGAACCTATCTATCATCTAGCAGAAGCCA ACGGCCTCCTGGAAGAGACAACCGATGGGGAACGCAGCGTGGGCCGCATCAGCCTCTA TTCCAAGAATGGCGTGGCCTGCTACCTTACCAACCACGGCCGAGGATCCCCAAGGAC GTGGTTGAGGAATTCAGCGATTTATACAACGAGGTCTATAACTTGACCCAGGAGTTCT TCCGGCACGATAAACCAGTCAATGCTGAAAGTCAAATAGCGTGGGGGTGTTACCCG AGAGGAGGTGCGTAACCGCATCAGGAATGACCCTGACGACCCAGAGGCTACCAAGCGC CTGAAGCTCGCCATGATCCAGCAGTACCTGAAGGTGGAGAGCTGTGAGAGCAGCTCAC ACAGCATGGACGAGGTGTCCCTGAGCGCCTTCGGGGAGTGGACCGAGATCCCCGGCGC TCACCACATCATCCCTCGGGCTTCATGCGGGTTGTGGAGCTGTGGCGGAGGGCATC CCTGCCACGTCATCCAGCTAGGGAAACCTGTCCGCTGCATTCACTGGGACCAGGCCT CAGCCCGCCCCAGAGGCCCTGAGATTGAGCCCCGGGTGTGCTAAAGAGGCAGTACAC CAGTTTCTTCGGCCAGGCCTGCCACAGAGAAGGTGGCTGCCATCCACCGCCTGGGC		

	ATTGGCACCACCGACAAGATCTTTCTGGAATTCGAGGAGCCCTTCTGGGGCCCTGAGT GCAACAGCCTACAGTTTGTGTGGGAGGACGAAGCGGAGAGCCACACCCTACCTACCC ACCTGAGCTCTGGTACCGCAAGATCTGCGGCTTTGATGTCTCTACCCGCTGAGCGC TACGGCCATGTGCTGAGCGGCTGGATCTGCGGGGAGGAGGCCCTCGTCATGGAGAAGT GTGATGACGAGGCAGTGGCCGAGATCTGCACGGAGATGCTGCGTCAGTTCACAGGGAA CCCCAACATTCCAAAACCTCGGCGAATCTTGCGCTCGGCCTGGGGCAGCAACCCTTAC TTCCGCGGCTCCTATTATACACGCAGGTGGGCTCCAGCGGGGCGGATGTGGAGAAGC TGGCCAAGCCCCTGCCGTACACGGAGAGCTCAAAGACAGCGCATGGAAGCTCCACAAA GCAGCAGCCTGGTCACCTTTTCTCTTCCAAGTGCCAGAACAGCCCCCTGGATGCTAAC AGGGGCGCCGTAAAGCCCATGCAGGTGCTGTTTTCCGGTGAGGCCACCCACCGCAAGT ACTATTCCACCACCCACGGTGCTCTGCTGTCCGGCCAGCGTGAGGCTGCCCGCCTCAT TGAGATGTACCGAGACCTCTTCCAGCAGGGGACCTGA		
	ORF Start: at 2		ORF Stop: TGA at 1601
	SEQ ID NO: 192	533 aa	MW at 59379.2kD
NOV13m, CG140122-04 Protein Sequence	TMQSCSSGDSADDP LSRGLRRRQPRVVVIGAGLAGLAAAKALLEQGFTDVTVLEAS SHIGGRVQSVKLGHATFELGATWIHGSHGNPIYHLAEANGLLEETDGRERSVGRISLY SKNGVACYLTNHGRRIPKDVVEEFS DLYNEVYNLTQEFFRHDKPVNAESQNSVGVFTR EEVRNRIRNDPDDPEATKRLKLAMIQQYLKVESCESSSHMSDEVSLSAFGEWTEIPGA HHIIPSGFMRVVELLAEGIPAHVIQLGKPVRCIHWDAQASARPRGPEIEPRGVLRQYT SFFRPGLPTEKVAATHRLGIGTTDKIFLEFEEFPWGPECNLSLQFVWEDEAESHTLTYP PELWYRKICGFDVLYPPERYGHVLSGWICGEEALVMEKCDDEAVAEICTEMLRQFTGN PNI PKPRRILRSAGWSNPYFRGSYSYTVQVSSGADVEKLAKPLPYTBSSKTAHGSSTK QQPGHLFSSKCPQLDANRGAVKPMQVLFSGEATHRKYYSTTHGALLSGQREARLI EMYRDLFQQGT		
	SEQ ID NO: 193	1513 bp	
NOV13n, CG140122-05 DNA Sequence	CACCATGCAAAGTTTGAATCCAGTGGTGACAGTGCGGATGACCCTCTCAGTCGCGGC CTACGGAGAAGGGGACAGCCTCGTGTGGTGGTGATCGGCGCCGGCTTGCTGGCCTGG CTGCAGCCAAAGCACTTCTTGAGCAGGGTTTACGGATGTCACTGTGCTTGAGGCTTC CAGCCACATCGGAGGCCGTGTGCAGAGTGTGAAACTTGACACAGCCACCTTTGAGCTG GGAGCCACCTGGATCCATGGCTCCCATGGGAACCTATCTATCATCTAGCAGAAGCCA ACGGCCTCCTGGAAGAGACAACCGATGGGGAACGACGCTGGGCGCATCAGCCTCTA TTCCAAGAATGGCGTGCCCTGTACCTTACCAACCACGCGCGAGGATCCCCAAGGAC GTGGTTGAGGAATTCAGCGATTTATACAACGAGGTCTATAACTTGACCCAGGAGTTCT TCCGGCAGGATAAACCAGTCAATGCTGAAAGTCAAAATAGCGTGCGGGGTGTTACCCG AGAGGAGGTGCGTAACCGCATCAGGAATGACCCTGACGACCCAGAGGCTACCAAGCGC CTGAAGCTCGCCATGATCCAGCAGTACCTGAAGGTGGAGAGCTGTGAGAGCAGCTCAC ACAGCATGGACGAGGTGTCCCTGAGCGCCTTCGGGGAGTGGAACCGAGATCCCCGGCC TCACCACATCATCCCTCGGGCTTCATGCGGGTTGTGGAGCTGCTGGCGGAGGGCATC CCTGCCACGTCATCCAGCTAGGGAACCTGTCCGCTGATTCACTGGGACAGGCTT CAGCCCGCCCCAGAGGCCCTGAGATTGAGCCCCGGGGTGCTAAAGAGGCAGTACAC CAGTTTCTTCCGGCCAGGCCCTGCCACAGAGAAGGTGGCTGCCATCCACCGCCTGGGC ATTGGCACCACCGACAAGATCTTTCTGGAATTCGAGGAGCCCTTCTGGGGCCCTGAGT GCAACAGCCTACAGTTTGTGTGGGAGGACGAAGCGGAGAGCCACACCCTACCTACCC ACCTGAGCTCTGGTACCGCAAGATCTGCGGCTTTGATGTCTCTACCCGCTGAGCGC TACGGCCATGTGCTGAGCGGCTGGATCTGCGGGGAGGAGGCCCTCGTCATGGAGAAGT GTGATGACGAGGCAGTGGCCGAGATCTGCACGGAGATGCTGCGTCAGTTCACAGGGAA CCCCAACATTCCAAAACCTCGGCGAATCTTGCGCTCGGCCTGGGGCAGCAACCCTTAC TTCCGCGGCTCCTATTATACACGCAGGTGGGCTCCAGCGGGCGGATGTGGAGAAGC TGGCCAAGCCCCTGCCGTACACGGAGAGCTCAAAGACAGCGCCCATGCAGGTGCTGTT TTCCGGTGAGGCCACCCACGCAAGTACTATTCCACCACCCACGGTGCTCTGCTGTCC GGCCAGCGTGAGGCTGCCCGCCTCATTGAGATGTACCGAGACCTCTTCCAGCAGGGGA CCTGA		
	ORF Start: at 2		ORF Stop: TGA at 1511
	SEQ ID NO: 194	503 aa	MW at 56191.7kD
NOV13n, CG140122-05	TMQSCSSGDSADDP LSRGLRRRQPRVVVIGAGLAGLAAAKALLEQGFTDVTVLEAS SHIGGRVQSVKLGHATFELGATWIHGSHGNPIYHLAEANGLLEETDGRERSVGRISLY SKNGVACYLTNHGRRIPKDVVEEFS DLYNEVYNLTQEFFRHDKPVNAESQNSVGVFTR		

Protein Sequence	EEVRNRIRNDPDDPEATKRLKRLAMIQQYLKVESCESSSSHSMDEVSLSAFGEWTEIPGAHHIIPSGFMRVVELLAEGIPAHVIQLGKPVRCIHWDQASARPRGPEIEPRGVLKRQYTSFFRPLPTEKVAIIHRLGIGTTDKIFLEFEEFPWGPECNSLQFVWEDEAESHTLTYPPELWYRKICGFDVLYPPERYGHLVSGWICGEEALVMEKCDDEAVAEICTEMLRQFTGNPNIPKPRRILRSAGWSNPYFRGSYSYTVGSSGADVEKLAKPLPYTESSKTAPMQVLFSGEATHRKYYSTTHGALLSGQREARLIEMYRDLFQQGT		
	SEQ ID NO: 195	1693 bp	
NOV13o, CG140122-06 DNA Sequence	CACCATGGGACATCATCACCACCATCACCAGTGTGTGAATCCAGTGGTGACAGTGGCGATGACCTCTCAGTCGCGGCCTACGGAGAAGGGGACAGCCTCGTGTGGTGGTATCGGCGCCGGCTTGGCTGGCTGGCTGCAGCCAAAGCACTTCTTGAGCAGGGTTTCACGGA TGTCACGTGTGCTTGAGGCTTCCAGCCACATCGGAGGCCGTGTGCAGAGTGTGAAACTTGGACACGCCACCTTTGAGCTGGGAGCCACCTGGATCCATGGCTCCCATGGGAACCTATCTATCATCTAGCAGAAGCCACGGCCTCCTGGAAGAGACAACCGATGGGGAACGCAGCGTGGGCGCATCAGCCTCTATTCCAAGATGGCGTGGCCTGCTACCTTACCAACCACGGCCGAGGATCCCAAGGACGTGGTTGAGGAATTCAGCGATTTATACAACGAGGTCTATAACTTGACCCAGGAGTTCTTCCGGCACGATAAACCACTCAATGCTGAAAGTCAAAATAGCGTGGGGGTGTTACCCGAGAGGAGGTGCGTAACCGCATCAGGAATGACCTGACGACCCAGAGGCTACCAAGCGCCTGAAGCTCGCCATGATCCAGCAGTACCTGAAGGTGGAGAGCTGTGAGAGCAGCTCACACAGCATGGACGAGGTGTCCTGAGCGCCTTCGGGGA GTGGAACGAGATCCCGGCGCTCACCACATCATCCCTCGGGCTTCATGCGGGTTGTGAGCTGCTGGCGGAGGGCATCCCTGCCACGTCATCCAGCTAGGGAACCTGTCCGCTGCATTCACTGGGACCAGGCCTCAGCCGCCCCAGAGGCCCTGAGATTGAGCCCCGGGTGAGGGCGACCAATCACGACACTGGGGAGGGTGGCCAGGGTGGAGAGGAGCCCCGGGGGGCAGGTGGGATGAGGATGAGCAGTGGTCGGTGGTGGTGGAGTGGCAGGACTGTGAGCTGATCCCGGCGGACCATGTGATTGTGACCGTGTGCTAGGTGTGCTAAAGAGGCA GTACACAGTTTCTTCCGGCCAGGCCTGCCACAGAGAAGGTGGCTGCCATCCACCGCCTGGGCATTGGCACCACGACAAGATCTTCTGGAATTCAGGAGGCCCTTCTGGGGCCCTGAGTGCAACAGCCTACAGTTTGTGTGGGAGGACGAAGCAGAGAGCCACACCCTCACCTACCCACCTGAGCTCTGGTACCGCAAGATCTGCGGCTTGATGTCTCTACCCGCCTGAGCGCTACGGCCATGTGCTGAGCGGCTGGATCTGCGGGGAGGAGGCCCTCGTCATGGAAGTGTGATGACGAGGCAGTGGCCGAGATCTGCACGGAGATGCTGCGTCAGTTCACAGGGAACCCCAACATTCCAAACCTCGGCGAATCTTGCGCTCGGCCCTGGGGCAGCAACCTTACTTCCGCGGCTCCTATTATACACGCAGGTGGGCTCCAGCGGGGCGGATGTGGAAGCTGGCCAAGCCCCCTGCGTACACGAGAGCTCAAAGACAGCGCCATGCAGGTGCTGTTTTCCGGTGAGGCCACCCACCGCAAGTACTATTCCACCACCCACGGTGCTCTGTGTCGGCCAGCGTGAGGCTGCCCCGCTCATTGAGATGTACCGAGACCTCTCCAGCAGGGGACCTGA		
	ORF Start: at 29		ORF Stop: TGA at 1691
	SEQ ID NO: 196	554 aa	MW at 61687.4kD
NOV13o, CG140122-06 Protein Sequence	QSCSSGSDSADDPLSRGLRRRQPRVVVIGAGLAGLAAAKALLEQGFDTVTVLEASSHIGGRVQSVKLGHATFELGATWIHGSHGNPIYHLAEANGLLEETDGGERSVGRISLYSKNGVACYLTNHGRRIPKDVVEEFSPLYNEVYNLTQEFRRHDKPVNAESQNSVGVFTREEVRNRIRNDPDDPEATKRLKRLAMIQQYLKVESCESSSSHSMDEVSLSAFGEWTEIPGAHHIIPSGFMRVVELLAEGIPAHVIQLGKPVRCIHWDQASARPRGPEIEPRGEGDHNHDTGEGGQGGEEPRGGRWDEDEQWSVVVECEDCELI PADHVI VTVSLGVLKRQYTSFFRPLPTEKVAIIHRLGIGTTDKIFLEFEEFPWGPECNSLQFVWEDEAESHTLTYPPELWYRKICGFDVLYPPERYGHLVSGWICGEEALVMEKCDDEAVAEICTEMLRQFTGNPNIPKPRRILRSAGWSNPYFRGSYSYTVGSSGADVEKLAKPLPYTESSKTAPMQVLFSGEATHRKYYSTTHGALLSGQREARLIEMYRDLFQQGT		
	SEQ ID NO: 197	1690 bp	
NOV13p, CG140122-07 DNA Sequence	CACCATGCAAAGTTGTGAATCCAGTGGTGACAGTGGGATGACCTCTCAGTCGCGGCCTACGGAGAAGGGGACAGCCTCGTGTGGTGGTATCGGCGCCGGCTTGGCTGGCCTGGCTGCAGCCAAAGCACTTCTTGAGCAGGGTTTACGGATGTCACTGTGCTTGAGCCTTCAGCCACATCGGAGGCCGTGTGCAGAGTGTGAACTTGGACACGCCACCTTTGAGCTGGAGCCACCTGGATCCATGGCTCCCATGGGAACCTATCTATCATCTAGCAGAAGCCAACGGCCTCCTGGAAGAGACAACCGATGGGGAACGCAGCGTGGGCGCATCAGCCTCTATTCCAAGATGGCGTGGCTGCTACCTTACCAACCACGGCCGAGGATCCCCAAGGAC		

	GTGGTTGAGGAATTCAGCGATTTATACAACGAGGTCTATAACTTGACCCAGGAGTTCT TCCGGCACGATAAACAGTCAATGCTGAAAGTCAAAATAGCGTGGGGGTGTTACCCG AGAGGAGGTGCGTAACCGCATCAGGAATGACCCTGACGACCCAGAGGCTACCAAGCGC CTGAAGCTCGCCATGATCCAGCAGTACCTGAAGGTGGAGAGCTGTGAGAGCAGCTCAC ACAGCATGGACGAGGTGTCCCTGAGCGCCTTCGGGGAGTGGACCGAGATCCCCGGCGC TCACCACATCATCCCTCGGGCTTCATGCGGGTTGTGGAGCTGCTGGCGGAGGGCATC CCTGCCCACGTATCCAGTAGGGAAACCTGTCCGCTGCATTACTGGGACCAGGCCT CAGCCCGCCCCAGAGGCCCTGAGATTGAGCCCCGGGGTGAGGGCGACCACAATCACGA CACTGGGGAGGGTGGCCAGGGTGGAGAGGAGCCCCGGGGGGGAGGTGGGATGAGGAT GAGCAGTGGTGGTGGTGGTGGAGTGCAGGACTGTGAGCTGATCCCGCGGACCATG TGATTGTGACCGTGTGCTAGGTGTGCTAAAGAGGCAGTACACCAAGTTTCTCCGGCC AGGCCTGCCCCACAGAGAAGGTGGCTGCCATCCACCGCTGGGCATTGGCACCACCGAC AAGATCTTTCTGGAATTCGAGGAGCCCTTCTGGGGCCCTGAGTGCAACAGCCTACAGT TTGTGTGGGAGGACGAAGCAGAGAGCCACACCCTCACCTACCCACCTGAGCTCTGGTA CCGCAAGATCTCGCGCTTTGATGTCCTTACCCGCTGAGCGCTACGGCCATGTGCTG AGCGGCTGGATCTGCGGGGAGGAGGCCCTCGTCATGGAGAAGTGTGATGACGAGGCAG TGGCCGAGATCTGCACGGAGATGCTGCGTCAGTTACAGGGAACCCCAACATTCCAAA ACCTCGGCGAATCTTGCCTCGGCCTGGGGCAGCAACCCCTACTTCCGCGGCTCCTAT TCATACACGCAGGTGGGCTCCAGCGGGGCGGATGTGGAGAAGCTGGCCAAGCCCCCTGC CGTACACGGAGAGCTCAAAGACAGCGCCATGCAGGTGCTGTTTTCCGGTGAGGCCAC CCACCGCAAGTACTATTCCACCACCCACGGTGCTCTGCTGTCCGGCCAGCGTGAGGCT GCCGCTCATTGAGATGTACCGAGACCTCTTCCAGCAGGGGACCCATCATCACCACC ATCACTGA		
	ORF Start: at 2		ORF Stop: TGA at 1688
	SEQ ID NO: 198	562 aa	MW at 62742.6kD
NOV13p, CG140122-07 Protein Sequence	TMQSCSSGDSADDPLSRGLRRRGQPRVVVIGAGLAGLAAAKALLEQGFTDVTVLEAS SHIGGRVQSVKLGHATFELGATWIHGSNGNPIYHLAEANGLLEETDGERSVGRISLY SKNGVACYLTNHGRRIPKDVVEEFSDLYNEVYNLTQEFFRHDKPVNAESQNSVGFTR EEVRNRIRNDPDDPEATKRLKLAMIQQYLKVESCESSSHSMDEVSLSAFGEWTEIPGA HHIIIPSGFMRVVELLAEGI PAHVILQKPVRCIHWDQASARPRGPEIEPRGEDHNDH TGEQGQGGEEPRGGRWDEDEQWSVVVECEDCELI PADHVIIVTVSLGVLKRYTSFFRP GLPTEKVAAIHRLGIGTTDKIFLEFEEFPWGPECNSLQFVWEDEAESHTLTYPPELWY RKICGFDVLYPPERYPYGHVLSGWI CGEEALVMEKCDDEAVAEICTEMLRQFTGNPNIPK PRILRSAWGSNPYFRGSYSYTVQVSSGADVEKLAKLPLPYTESSTKAPMQVLFSGEAT HRKYYSTTHGALLSGQREARLIEMYRDLFQQGTHHHHHH		
	SEQ ID NO: 199	1680 bp	
NOV13q, CG140122-08 DNA Sequence	TCCACCATGCAAAGTTGTGAATCCAGTGGTGACAGTGCGGATGACCCTCTCAGTCGCG GCCTACGGAGAAGGGGACAGCCTCGTGTGGTGGTGTGATCGGCGCCGGCTTGGCTGGCCT GGCTGCAGCCAAAGCACTTCTTGAGCAGGGTTTACCGATGTCACTGTGCTTGAGGCT TCCAGCCACATCGGAGGCGGTGTGCAGAGTGTGAACTTGGACACGCCACCTTGTAGC TGGGAGCCACCTGGATCCATGGCTCCCATGGGAACCTATCTATCATCTAGCAGAAGC CAACGGCCTCCTGGAAGAGACAACCGATGGGGAACGCAGCGTGGGCCGCATCAGCCTC TATTCCAAGAAATGGCGTGGCCTGCTACCTTACCAACCAAGGCCGAGGATCCCCAAGG ACGTGGTTGAGGAATTCAGCGATTTATACAACGAGGTCTATAACTTGACCCAGGAGTT CTTCGGGCACGATAAACAGTCAATGCTGAAAGTCAAAATAGCGTGGGGGTGTTACCC CGAGAGGAGGTGCGTAACCGCATCAGGAATGACCCTGACGACCCAGAGGCTACCAAGC GCCTGAAGCTCGCCATGATCCAGCAGTACCTGAAGGTGGAGAGCTGTGAGAGCAGCTC ACACAGCATGGACGAGGTGTCCCTGAGCGCCTTCGGGGAGTGGACCGAGATCCCCGGC GCTCACCACATCATCCCTCGGGCTTCATGCGGGTTGTGGAGCTGCTGGCGGAGGGCA TCCCTGCCCACGTATCCAGCTAGGGAAACCTGTCCGCTGCATTACTGGGACCAGGC CTCAGCCCCCCCCAGAGGCCCTGAGATTGAGCCCCGGGGTGAGGGCGACCACAATCAC GACTGAGGAGGGTGGCCAGGGTGGAGAGGAGCCCCGGGGGGGAGGTGGGATGAGG ATGAGCAGTGGTGGTGGTGGTGGAGTGCAGGAGTGTGAGCTGATCCCGCGGACCA TGTGATTGTGACCGTGTGCTAGGTGTGCTAAAGAGGCAGTACACAGATTCTTCCGG CCAGGCCTGCCACAGAGAAGGTGGCTGCCATCCACCGCCTGGGCATTGGCACCACCG ACAAGATCTTCTGGAATTCGAGGAGCCCTTCTGGGGCCCTGAGTGCAACAGCCTACA GTTTGTGTGGGAGGACGAAGCAGAGAGCCACACCCTCACCTACCCACCTGAGCTCTGG TACCGCAAGATCTGCGGCTTTGATGTCCTTACCCGCTGAGCGCTACGGCCATGTGC		

	TGAGCGGCTGGATCTGCGGGGAGGAGGCCCTCGTCATGGAGAAGTGTGATGACGAGGC AGTGGCCGAGATCTGCACGGAGATGCTGCGTCAGTTACAGGGAACCCCAACATTCCA AAACCTCGGCGAATCTTGCGCTCGGCCTGGGGCAGCAACCCCTACTTCCGCGGCTCCT ATTATACACGCAGGTGGGCTCCAGCGGGGCGGATGTGGAGAAGCTGGCCAAGCCCT GCCGTACACGGAGAGCTCAAAGACAGCGCCCATGCAGGTGCTGTTTTCCGGTGAGGCC ACCCACCGCAAGTACTATTCCACCAACCGGTGCTCTGCTGTCCGGCCAGCGTGAGG CTGCCCGCCTCATTGAGATGTACCGAGACCTCTCCAGCAGGGGACCTGAAAGCTT		
	ORF Start: at 1		ORF Stop: TGA at 1672
	SEQ ID NO: 200	557 aa	MW at 62006.8kD
NOV13q, CG140122-08 Protein Sequence	STMQSCSSGDSADDPLSRGLRRRQPRVVIGAGLAGLAAAKALLEQGFTDVTVLEA SSHIGGRVQSVKLGHATFELGATWIHGSHGNPIYHLAEANGLLEETDGERSVGRISL YSKNGVACYLTNHGRRIPKDVVEEFSDLYNEVYNLTQEFRRHDKPVNAESQNSVGVT REEVRNRIRNDPDDPEATKRLKLAMIQQYLKVESCESSSHSMDEVSLSAFGEWTEIPG AHHIIPSGFMRVVELLAEGIPAHVIQLGKPVRCIHWDAQASARPRGPBIEPRGEDHND DTGEGGQGGEPRGGRWDEDEQWSVVVECEDCELI PADHVI VTVSLGLKRQYTSFFR PGLPTEKVAAIHRLGIGTTDKIFLEFEEPFWGPECNSLQFVWEDEAESHTLTYPPELW YRKICGFDVLYPPERYGHVLSGWI CGEEALVMEKCDDEAIAEICTEMLRQFTGNPNIP KPRRILRSAWGSNPYFRGSYSYTQVSSGADVEKLAKPLPYTESSKTAPMQVLFSGEA THRKYYSTTHGALLSGQREARLIEMYRDLFQQGT		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 13B.

Table 13B. Comparison of NOV13a against NOV13b through NOV13q.		
Protein Sequence	NOV13a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV13b	1..280	280/280 (100%)
	2..281	280/280 (100%)
NOV13c	1..555	502/585 (85%)
	2..533	502/585 (85%)
NOV13d	2..555	553/554 (99%)
	10..563	553/554 (99%)
NOV13e	1..555	554/555 (99%)
	2..556	554/555 (99%)
NOV13f	1..555	554/555 (99%)
	2..556	554/555 (99%)
NOV13g	1..555	554/555 (99%)
	8..562	554/555 (99%)
NOV13h	1..555	554/555 (99%)
	2..556	554/555 (99%)
NOV13i	1..555	554/555 (99%)
	2..556	554/555 (99%)
NOV13j	2..555	553/554 (99%)
	10..563	553/554 (99%)
NOV13k	1..555	502/555 (90%)

	1..502	502/555 (90%)
NOV13l	1..280 2..281	280/280 (100%) 280/280 (100%)
NOV13m	1..555 2..533	502/585 (85%) 502/585 (85%)
NOV13n	1..555 2..503	502/555 (90%) 502/555 (90%)
NOV13o	2..555 1..554	553/554 (99%) 553/554 (99%)
NOV13p	1..555 2..556	554/555 (99%) 554/555 (99%)
NOV13q	1..555 3..557	554/555 (99%) 554/555 (99%)

Further analysis of the NOV13a protein yielded the following properties shown in Table 13C.

Table 13C. Protein Sequence Properties NOV13a	
PSort analysis:	0.7900 probability located in plasma membrane; 0.4802 probability located in microbody (peroxisome); 0.3000 probability located in Golgi body; 0.2000 probability located in endoplasmic reticulum (membrane)
SignalP analysis:	Cleavage site between residues 41 and 42

5. A search of the NOV13a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 13D.

Table 13D. Geneseq Results for NOV13a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV13a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAB73670	Human oxidoreductase protein ORP-3 - Homo sapiens, 555 aa. [WO200144448-A2, 21-JUN-2001]	1..555 1..555	554/555 (99%) 554/555 (99%)	0.0
AAB12164	Hydrophobic domain protein from clone HP10673 isolated from Thymus cells - Homo sapiens, 555 aa. [WO200029448-A2, 25-MAY-2000]	1..555 1..555	554/555 (99%) 554/555 (99%)	0.0

AAM79546	Human protein SEQ ID NO 3192 - Homo sapiens, 518 aa. [WO200157190-A2, 09-AUG-2001]	1..510 7..516	508/510 (99%) 508/510 (99%)	0.0
AAM78562	Human protein SEQ ID NO 1224 - Homo sapiens, 513 aa. [WO200157190-A2, 09-AUG-2001]	1..510 1..511	501/511 (98%) 501/511 (98%)	0.0
AAU21643	Novel human neoplastic disease associated polypeptide #76 - Homo sapiens, 335 aa. [WO200155163-A1, 02-AUG-2001]	273..555 53..335	282/283 (99%) 282/283 (99%)	e-171

In a BLAST search of public sequence databases, the NOV13a protein was found to have homology to the proteins shown in the BLASTP data in Table 13E.

Table 13E. Public BLASTP Results for NOV13a				
Protein Accession Number	Protein/Organism/Length	NOV13a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q96QT3	Polyamine oxidase isoform-1 - Homo sapiens (Human), 555 aa.	1..555 1..555	555/555 (100%) 555/555 (100%)	0.0
Q9NWM0	CDNA FLJ20746 fis, clone HEP06040 - Homo sapiens (Human), 555 aa.	1..555 1..555	554/555 (99%) 554/555 (99%)	0.0
Q99K82	Similar to hypothetical protein - Mus musculus (Mouse), 555 aa.	1..554 1..554	528/554 (95%) 537/554 (96%)	0.0
Q9NP51	DJ779E11.1.5 (Novel flavin containing amine oxidase (Translation of cDNA DFKZp761P0724 (Em:AL162058)) (Isoform 5)) - Homo sapiens (Human), 412 aa (fragment).	144..555 1..412	411/412 (99%) 411/412 (99%)	0.0
Q9H6H1	CDNA: FLJ22285 fis, clone HRC03956 - Homo sapiens (Human), 389 aa.	197..555 1..389	357/389 (91%) 357/389 (91%)	0.0

PFam analysis predicts that the NOV13a protein contains the domains shown in the Table 13F.

Table 13F. Domain Analysis of NOV13a			
Pfam Domain	NOV13a Match Region	Identities/ Similarities for the Matched Region	Expect Value

FAD_binding_3	27..141	24/142 (17%) 74/142 (52%)	0.31
Amino_oxidase	34..544	124/574 (22%) 366/574 (64%)	1.8e-28

**Example 14.**

The NOV14 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 14A.

Table 14A. NOV14 Sequence Analysis			
	SEQ ID NO: 201	2058 bp	
NOV14a, CG140316-01 DNA Sequence	ATGGAGCCCGAAGCCCCCGTCGCCGCCACCCCATCAGCGCGGCTACCTGCTGACAC GGAACCCCTCACCCTCAACAGGACTTGGCCTTTACCCTGGAAGAGAGACAGCAATTGAA CATTTCATGGATTGTTGCCACCTTCCCTTCAACAGTCAGGAGATCCAGGTTCTTAGAGTA GTAAAAAATTCGAGCATCTGAACCTCTGACTTTGACAGGTATCTTCTCTTAATGGATC TCCAAGATAGAAATGAAAACTCTTTTATAGAGTGCTGACATCTGACATTGAGAAATT CATGCCTATTGTTTATACTCCCACTGTGGGTCTGGCTTGCCAACAATATAGTTTGGTG TTTTCGGAAGCCAAGAGGTCTCTTTATTACTATCCACGATCGAGGCATATTGCTTCAG TTCTCAATGCATGGCCAGAAGATGTCATCAAGGCCATTGTGGTGACTGATGGAGAGCG TATTCTTGGCTTGGGAGACCTTGGCTGTAATGGAATGGGCATCCCTGTGGGTAAATTG GCTCTATATACAGCTTGGGAGGGATGAATCCTCAAGAATGTCTGCCTGTCTCTGG ATGTGGGAACCGAAAATGAGGAGTTACTTAAAGATCCACTCTACATTGGACTACGGCA GAGAAGAGTAAGAGGTTCTGAATATGATGATTTTTTGGACGAATTCATGGAGGCAAGT TCTTCCAAGTATGGCATGAATTGCCTTATTCAGTTTGAAGATTTTGCCAATGTGAATG CATTTGCTCTCCTGAACAAGTATCGAAACAGTATGCACATTCAATGATGATTTCA AGGAACAGCATCTGTTGCAGTTGCAGGTCTCCTTGCAGCTCTTCAATAACCAAGAAC AAAGTGTCTGATCAACAATACTATTCCAAGGAGCTGGAGAGGCTGCCCTAGGGATTG CACACCTGATGTGATGGCCTTGGAAAAAGAAGGTTTACCAAAAGAGAAAGCCATCAA AAAGATATGGCTGGTTGATTCAAAAGGATTAATAGTTAAGGGACGTGCTTCTCTAACA CAAGAGAAAGAGAAGTTTGCCCATGAACATGAAGAAATGAAGAACCTAGAAGCCATTG TTCAAGAAATAAAACCAACTGCCCTCATAGGAGTTGCTGCAATTTGGTGGTGATTCTC AGAACAAATTCTCAAGATATGGCTGCCTTCAATGAACGGCCTATTATTTTGGCTTTG AGTAATCCAACTAGCAAAGCAGAATGTTCTGCAGAGCAGTGCTACAAAATAACCAAGG GACGTGCAATTTTTGGCAGTGGCAGTCCCTTTTGATCCAGTCACTCTTCCAAATGGACA GACCTATATCCTGGCCAAGGCAACAATTCCTACGTGTTCCCTGGAGTTGCTCTTGGT GTTGTGGCGTGTGGATTGAGGCAGATCACAGATAATTTTCTCACTACTGCTGAGG TTATAGCTCAGCAAGTGTGAGATAAACACTTGGAAAGAGGTCGGCTTTATCCTCCTTT GAATACCATTAGAGATGTTTCTCTGAAAATTGCAGAAAAGATTGTGAAAGATGCATAC CAAGAAAAGACAGCCACAGTTTATCCTGAACCGCAAAACAAAGAAGCATTGTGCCGCT CCCAGATGTATAGTACTGATTATGACCAGATTCTACCTGATTGTTATTCTTGGCCTGA AGAGGTGCAGAAAATACAGACCAAGTTGACCAGTAGGATAATAGCAAAACATTTCTAA CTCTATTAAATGAGGTCTTTAAACCTTTTATAATTTTAAAGTTTGAATCTTTTATAA TGATTCTAAGACACTTAGATTAAAGATTTTACTTTAACAGTCTAAAAATTGATAGAAG AATATCGATATAAATTGGGATAAACATCACATGAGACAATTTTGCTTCACTTTGCCTT CTGGTTATTATGGTTTCTGTCTGAATTATTCTGCCTACGTTCTCTTTAAAGCTGTT GTACGTACTACGGAGAACTCATCATTTTATACAGGACACTAATGGGAAGACCAAAA TTACTAATAAATTGAAATAACCAACATT		
	ORF Start: ATG at 1		ORF Stop: TAG at 1717
	SEQ ID NO: 202	572 aa	MW at 64148.9kD
NOV14a, CG140316-01 Protein Sequence	MEPEAPRRRHQYLLTRNPHLNKDLAFTLEERQQLNIHGLLPPSFNSQEIQVLRV VKNFEHLNSDFDRYLLMLDQDRNEKLFYRVLTSDIEKFMPVYVTPVGLACQQYSLV FRKPRGLFITIHDRGHIASVLNAPEDVIKAIVVTDGERILGLDGLCGNGMGPVQGL ALYTACGMNPNQELPVILDVGTENEELLKDPYIGLRQRRVRGSEYDDFLDEFMEAV		

	SSKYGMNCLIQFEDFANVNAFRLLNKYRNOYCTFNDDIQGTASVAVAGLLAALRITKN KLSDQITILFQGAGEAALGIAHLIVMALEKEGLPKEKAIKKIWLVDKGLIVKGRASLT QEKEKFAHEHEEMKNLEAIVQEI KPTALIGVAAIGGAFSEQILKDMAAFNERPIIFAL SNPTSKAECSAEQCYKITKGRAIFASGSPFDPVTL PNGQTLYPGQGNNSYVFPGVALG VVACGLRQITDNI FLTTAEVIAQQVSDKHLEEGRLYPPLNTIRDVSLKIAEKIVKDAY QEKATATVYPEPQNKEAFVRSQMYSTDYDQILPDCYSWPEEVQKIQTQKVDQ		
	SEQ ID NO: 203	2058 bp	
NOV14b, CG140316-01 DNA Sequence	ATGGAGCCCGAAGCCCCCGTCGCCGCCACACCCATCAGCGCGGCTACCTGCTGACAC GGAACCCTCACCTCAACAAGGACTTGGCCTTTACCCTGGAAGAGAGACAGCAATTGAA CATTTCATGGATTGTTGCCACCTTCCCTCAACAGTCAGGAGATCCAGGTTCTTAGAGTA GTAAAAAATTTTCAGCATCTGAACCTGACTTTGACAGGTATCTTCTCTTAATGGATC TCCAAGATAGAAATGAAAACTCTTTTATAGAGTGCTGACATCTGACATTGAGAAATT CATGCCTATTGTTTATACTCCCACTGTGGGTCTGGCTTGCCAACAATATAGTTGGTG TTTCGGAAGCCAAGAGGTCTCTTTATTACTATCCACGATCGAGGGCATATTGCTTCAG TTCTCAATGCATGGCCAGAAGATGTCATCAAGGCCATTGTGGTGACTGATGGAGAGCG TATTCTTGGCTTGGGAGACCTTGGCTGTAATGGAATGGGCATCCCTGTGGGTAAATTG GCTCTATATACAGCTTGGCGAGGGATGAATCCTCAAGAATGTCTGCCTGTCTATTCTGG ATGTGGGAACCGAAAAATGAGGAGTTACTTAAAGATCCACTCTACATTGGACTACGGCA GAGAAGAGTAAGAGGTTCTGAATATGATGATTTTGTGGACGAATTCATGGAGGCAGTT TCTTCCAAGTATGGCATGAATTGCCTTATTTCAGTTTGAAGATTTTGCCAATGTGAATG CATTTCGTCTCCTGAACAAGTATCGAAACCAGTATTGCACATTCATATGATGATATTCA AGGAACAGCATCTGTTGCAGTTGCAGGTCTCCTTGACGCTCTTCGAATAACCAAGAAC AACTGTCTGATCAACAATACTATTCCAAGGAGCTGGAGAGGCTGCCCTAGGGATTG CACACCTGATTGTGATGGCCTTGGAAAAAGAAGGTTTACCAAAAGAGAAAGCCATCAA AAAGATATGGCTGGTTGATTCAAAAGGATTAATAGTTAAGGGACGTGCTTCCTTAACA CAAGAGAAAGAGAAGTTTGCCCATGAACATGAAGAAATGAAGAACCCTAGAAGCCATTG TTCAAGAAATAAAACCAACTGCCCTCATAGGAGTTGCTGCAATTGGTGGTGCAATTCTC AGAACAAATTTCTCAAGATATGGCTGCCTTCAATGAACGGCTATTATTTTGGCTTTG AGTAATCCAAGTACGAAAGCAGAAATGTTCTGCAGAGCAGTGCTACAAAATAACCAAGG GACGTGCAATTTTGGCAGTGGCAGTCTCTTTGATCCAGTCACTCTTCCAAATGGACA GACCTATATCCTGGCCAAGGCAACAATTCTACGTGTTCCCTGGAGTTGCTCTTGGT GTTGTGGCGTGTGGATTGAGGCAGATCACAGATAATATTTTCTCCTACTCTGCTGAGG TTATAGCTCAGCAAGTGTGAGATAAACTTGAAGAGGGTGGCTTTATCCTCCTTT GAATACCATTAGAGATGTTTCTTGAAAATTGCAGAAAAGATTGTGAAAGATGCATAC CAAGAAAAGACAGCCACAGTTTATCCTGAACCGCAAAACAAAGAAGCAATTGTCCGCT CCCAGATGTATAGTACTGATTATGACCAGATTCTACCTGATTGTTATCTTGGCCTGA AGAGGTGCAGAAAATACAGACCAGGTTGACAGTAGGATAATAGCAAAACATTTCTAA CTCTATTAATGAGGTCTTTAAACCTTTTATAATTTTAAAGGTTGGAATCTTTTATAA TGATTCTAAGACACTTAGATTAGATTTTACTTTAACAGTCTAAAAATTGATAGAAG AATATCGATATAAATTTGGGATAAATCATCATGAGACAATTTTGCTTCACTTTGCCTT CTGGTTATTTATGGTTTCTGTCTGAATTATTCTGCCTACGTTCTCTTTAAAGCTGTT GTACGTACTACGGAGAACTCATATTTTATACAGGACACTAATGGGAAGACCAAAA TTACTAATAAATTGAAATAACCAACATT		
	ORF Start: ATG at 1		ORF Stop: TAG at 1717.
	SEQ ID NO: 204	572 aa	MW at 64148.9kD
NOV14b, CG140316-01 Protein Sequence	MEPEAPRRRHTHQRYLLTRNPHLNKDLAFTLEERQQLNIHGLLPSPFNSQEIQVLRV VKNFEHLNSDFDRYLLMLDQDRNEKLFYRVLTSDIEKFMPIVYTPTVGLACQQYSLV FRKPRGLFITIHDRGHIAVLNAPEDVIKAIIVTDGERILGLDGLCNGMGIPVGLK ALYTACGMNPQCELPVILDVGTENEELLKDPLYIGLRQRRVRGSEYDDFLDEFMEAV SSKYGMNCLIQFEDFANVNAFRLLNKYRNOYCTFNDDIQGTASVAVAGLLAALRITKN KLSDQITILFQGAGEAALGIAHLIVMALEKEGLPKEKAIKKIWLVDKGLIVKGRASLT QEKEKFAHEHEEMKNLEAIVQEI KPTALIGVAAIGGAFSEQILKDMAAFNERPIIFAL SNPTSKAECSAEQCYKITKGRAIFASGSPFDPVTL PNGQTLYPGQGNNSYVFPGVALG VVACGLRQITDNI FLTTAEVIAQQVSDKHLEEGRLYPPLNTIRDVSLKIAEKIVKDAY QEKATATVYPEPQNKEAFVRSQMYSTDYDQILPDCYSWPEEVQKIQTQKVDQ		
	SEQ ID NO: 205	1750 bp	
NOV14c,	CGCGGATCCATGGAGCCCGAAGCCCCCGTCGCCGCCACACCCATCAGCGCGGCTACC		

254047949 DNA Sequence	TGCTGACACGGAACCCCTCACCTCAACAAGGACTTGGCCTTTACCCTGGAAGAGAGACA GCAATTGAACATTTCATGGATTGTTGCCACCTTCCTTCAACAGTCAGGAGATCCAGGTT CTTAGAGTAGTAAAAAATTTTCGAGCATCTGAACTCTGACTTTGACAGGTATCTTCTCT TAATGGATCTCCAAGATAGAAATGAAAACTCTTTATAGAGTGCTGACATCTGACAT TGAGAAATTCATGCCATTGTTTATACTCCCACTGTGGGTCTGGCTTGCCAACAATAT AGTTTGGTGTTCGGAAGCCAAGAGGTCTCTTTATTACTATCCAGATCGAGGGCATA TTGCTTCAGTTCTCAATGCATGGCCAGAAGATGTCATCAAGGCCATTGTGGTGACTGA TGGAGAGCGTATTCTTGGCTTGGGAGACCTTGGCTGTAATGGAATGGGCATCCCTGTG GGTAAATTGGCTCTATATACAGCTTGCAGGAGGATGAATCCCAAGAATGTCTGCCTG TCATTCTGGATGTGGGAACCGAAAATGAGGAGTTACTTAAAGATCCACTCTACATTGG ACTACGGCAGAGAAGAGTAAGAGGTCTGAATATGATGATTTTGGACGAATTCATG GAGGCAGTTCTTCCAAGTATGGCATGAATTGCCTTATTCAGTTTGAAGATTTTGCCA ATGTGAATGCATTTCTGCTCTCTGAACAAGTATCGAAACCAGTATTGCACATTCAATGA TGATATTCAAGGAACAGCATCTGTTGCAGTTGCAGGTCTCCTTGCAGCTCTTCAATA ACCAAGAACAACTGTCTGATCAACAATACTATTCCAAGGAGCTGGGGAGGCTGCC TAGGGATTGCACACCTGATTGTGATGGCCTTGAAAAAGAAGGTTTACCAAAAGAGAA AGCCATCAAAAAGATATGGCTGGTTGATTCAAAAGGATTAATAGTTAAGGACGTGCT TCCTTAACACAAGAGAAAAGAGAAGTTGCCCATGAACATGAAGAAATGAAGAACCTAG AAGCCATTGTTCAAGAAATAAAACCAACTGCCCTCATAGGAGTTGCTGCAATTGGTGG TGCATTCTCAGAACAAATTCTCAAGATATGGCTGCCTTCAATGAACGGCTATTATT TTTGCTTTGAGTAATCCAAGTACGAAAGCAGAAATGTTCTGCAGAGCAGTGCTACAAAA TAACCAAGGACGTGCAATTTTGGCAGTGGCAGTCTCTTTGATCCAGTCACTCTTCC AAATGGACAGACCCATATCTTGGCCAAGGCAACAATTCCATGTGTTCCCTGGAGTT GCTCTTGGTGTGTGGCGTGTGGATTGAGGCAGATCACAGATAATATTTCTCTACTA CTGCTGAGGTTATAGCTCAGCAAGTGTGAGATAAACACTTGAAGAGGGTCTGGCTTTA TCCTCCTTTGAATACCATTAGAGATGTTCTCTGAAAATTGCAGAAAAGATTGTGAAA GATGCATACCAAGAAAAGACAGCCACAGTTTATCCTGAACCGCAAAACAAAGAAGCAT TTGTCCGCTCCAGATGTATAGTACTGATTATGACCAGATTCTACCTGATTGTTATTCT TTGGCCTGAAGAGGTGCAGAAAATACAGACCAAGTTGACCAGTAGGGTGGCGGCCG TTTTTTCCTT		
	ORF Start: at 1		ORF Stop: TAG at 1726
	SEQ ID NO: 206	575 aa	MW at 64449.2kD
NOV14c, 254047949 Protein Sequence	RGSMEPEAPRRRHQRYLLTRNPHLNKDLAFTLEERQQLNIHGLLPPSFNSQEIQV LRVVKNFHNLNSDFRYLLMDLQDRNEKLFYRVLTSDIEKFMPIVYPTVGLACQQY SLVFRKPRGLFITIHDRGHIAVLNAPEDVIKAIIVTDERILGLDGLGNGMGIPIV GKLALYTACGGMNPQECPLVILDVGTENEELLKDPYIIGLRQRRVRGSEYDDFLDEFM EAVSSKYMGNCLIQFEDFANVNAFRLNKNRYNYCTFNDDIQGTASVAVAGLLAALRI TKNKLSDQTLIFQAGAEALGIAHLIVMALEKEGLPKKAIKKIWLVDKGLIVKGR SLTQEKEKFAHEHEEMKNLEAIVQEIKPTALIGVAAIGGAFSEQILKDMAAFNERPII FALSNPTSKECSAEQCYKITKGRAIFASGSPFDPVTLNPGQTLYPGQGNNSYVFPV ALGVVACGLRQITDNIFLTAEVIAQQVSDKHLEEGRLYPPLNTIRDVSLKIAEKIVK DAYQEKATATVYPEPQNEAFVRSQMYSTDYDQILPDCYSWPBEVQKIQTQKVDQ		
	SEQ ID NO: 207	1752 bp	
NOV14d, 258280122 DNA Sequence	AGCCCGAAGCCCCCGTCGCCGCCACCCATCAGCGCGGCTACCTGCTGACACGGAA CCCTCACCTCAACAAGGACTTGGCCTTTACCCTGGAAGAGAGACAGCAATTGAACATT CATGGATTGTTGCCACCTTCCTTCAACAGTCAGGAGATCCAGGTTCTTAGAGTAGTAA AAAATTTGAGCATCTGAACTCTGACTTTGACAGGTATCTTCTTAAATGGATCTCCA AGATAGAAATGAAAACTCTTTTATAGAGTGCTGACATCTGACAAATTCATG CCTATTGTTTATACTCCCACTGTGGGTCTGGCTTGCCAACAATATAGTTTGGTGTTC GGAAGCCAAGAGGTCTCTTTATTACTATCCACGATCGAGGGCATATTGCTTCAGTTCT CAATGCATGGCCAGAAGATGTCATCAAGGCCATTGTGGTGACTGATGGAGAGCGTATT CTTGGCTTGGGAGACCTTGGCTGTAATGGAATGGGCATCCCTGTGGGTAATTTGGCTC TATATACAGCTTGGGAGGGATGAATCCCAAGAATGTCTGGCTGCTCATTCTGGATGT GGGAACCGAAAAATGAGGAGTTACTTAAAGATCCACTCTACATTGGACTACGGCAGAGA AGAGTAAGAGGTTCTGAATATGATGATTTTTTGGACGAATTCATGGAGGCAGTTTCTT CCAAGTATGGCATGAATGCCTTATTCAGTTTGAAGATTTTGCCAATGTGAATGCATT TCGTCTCCTGAACAAGTATCGAAACCAGTATGCACATTCAATGATGATATTCAAGGA ACAGCATCTGTTGCAGTTGCAGGTCTCCTTGACGCTCTTCAATAACCAAGAACAAAC		

	<p>TGCTCTGATCAAACAATACTATTTCCAAGGAGCTGGGGAGGCTGCCCTAGGGATTGCACA  CCTGATTGTGATGGCCTTGGAAAAAGAAGGTTTACCAAAAGAGAAAGCCATCAAAAAG  ATATGGCTGGTTGATTCAAAGGATTAATAGTTAAGGGACGTGCTTCCTTAACACAAG  AGAAAGAGAAAGTTTCCCCATGAACATGAAGAAATGAAGAACCTAGAAGCCATTGTTCA  AGAAATAAAACCAACTGCCCTCATAGGAGTTGCTGCAATTGGTGGTGCAATTCTCAGAA  CAAATTCTCAAAGATATGGCTGCCTTCAATGAACGGCCTATTATTTTGGCTTTGAGTA  ATCCAAGTAGCAAAGCAGAAATGTTCTGCAGAGCAGTGCTACAAAATAACCAAGGGACG  TGCAATTTTGGCAGTGGCAGTCTTTTGTATCCAGTCACTCTTCAAATGGACAGACC  CTATATCCTGGCCAAGGCAACAATTCCTATGTGTTCCCTGGAGTTGCTCTTGGTGTG  TGGCGTGTGGATTGAGGCAGATCAGATAAATATTTTCTCACTACTGCTGAGGTTAT  AGCTCAGCAAGTGTGAGATAAACACTTGGAGAGGGTCGGCTTTATCCTCCTTTGAAT  ACCATTAGAGATGTTTCTCTGAAAATTGCAGAAAAGATTGTGAAGATGCATACCAAG  AAAAGACAGCCACAGTTTATCCTGAACCGCAAAACAAAGAAGCATTGTCCGCTCCCA  GATGTATAGTACTGATTATGACAGATTCTACCTGATTGTTATTCTTGGCCTGAAGAG  GTGCAGAAAATACAGACCAAGTTGACCAAGTAGGGTGGCGGCCGCACTCGAGCACCAC  CACCACCACCAC</p>		
	ORF Start: at 3		ORF Stop: TAG at 1713
	SEQ ID NO: 208	570 aa	MW at 63888.6kD
NOV14d, 258280122 Protein Sequence	<p>PEAPRRRHTHQRYLLTRNPHLNKDLAFTLEERQQLNIHGLLPPSFNSQEIQVLRVVK  NFEHLNSDFDRYLLMLDLQDRNEKLFYRVLTSDIEKFMPIVYTPTVGLACQQYSLVFR  KPRGLFITIHDRGHIAVLNAPEDVIKAI VVT DGERILGLDGLCGNGMGI PVGKLAL  YTACGGMNPQECPLVILDVGTENEELLKDPLYIGLRQRRVRGSEYDDFLDEFMEAVSS  KYGMNCLIQFEDFANVNAFRLNKNYRNQYCTFNDDIQGTASVAVAGLLAALRITKNKL  SDQTILFQGAGEAALGIAHLIVMALEKEGLPKKAIKKIWLVD SKGLIVKGRASLTQE  KEKFAHEHEEMKNLEAIVQEI KPTALIGVAAIGGAFSEQILKDMAAFNERPI I FALSN  PTSKAECSEAQQYKIKTGRAIFASGSPFPDPTLPNGQTLYPGQGNNSYVFPGVALGVV  ACGLRQITDNIFLTAEVIAQQVSDKHLEEGRLYPPLNTIRDVSLKIAEKIVKDAYQE  KTATVYPEPQNKFAFVRSQMYSTDYDQILPDCYSWPPEVQKIQTQKVDQ</p>		
	SEQ ID NO: 209	1743 bp	
NOV14e, 258330149 DNA Sequence	<p>ACCATGGGCCACCATCACCACCATCAGAGCCCCGAAGCCCCCGTCGCCGCCACACCC  ATCAGCGCGGCTACCTGCTGACACGGAACCTCACCTCAACAAGGACTTGGCCTTTAC  CCTGGAAGAGAGACAGCAATTGAAACATTATGGATTGTGTCACCTTCCTTCAACAGT  CAGGAGATCCAGGTTCTTAGAGTAGTAAAAAATTTGAGCATCTGAACTCTGACTTTG  ACAGGTATCTTCTTAAATGATCTCCAAGATAGAAATGAAAACTCTTTTATAGAGT  GCTGACATCTGACATTGAGAAATTCATGCCTATTGTTTATACTCCCACTGTGGGTCTG  GCTTGCCAACAATATAGTTTGGTGTTCGGAAGCCAAGAGGCTCTTTATTACTATCTC  ACGATCGAGGGCATATTGCTTCAGTTCTCAATGCATGGCCAGAGATGTCATCAAGGC  CATTGTGGTGAAGTATGAGAGCGTATTCTTGGCTTGGGAGACCTTGGCTGTAATGGA  ATGGGCATCCCTGTGGGTAAATTGGCTCTATATACAGCTTGCAGAGGATGAATCCTC  AAGAATGTCTGCTGTCTTCTGGATGTGGGAACCGAAAATGAGGAGTTACTTAAAGA  TCCACTCTACATTGGAATACGGCAGAGAAGAGTAAGAGGTTCTGAATATGATGATTTT  TTGGACGAATTATGAGGAGGATTTCTTCCAAGTATGGCATGAATTGCCTTATTCAGT  TTGAAGATTTTCCAATGTGAATGCATTTCGTCTCCTGAACAAGTATCGAAACAGTA  TTGCACATTCAATGATGATATTCAAGGAACAGCATCTGTTGCAGTTGCAGGTCTCCTT  GCAGCTCTTCAATAAACCAGAACAACTGTCTGATCAACAATACTATTCCAAGGAG  CTGGGGAGGCTGCCCTAGGGATTGCACACCTGATTGTGATGGCCTTGGAAAAAGAAGG  TTTACCAAAAGAGAAAGCCATCAAAAAGATATGGCTGGTTGATTTCAAAGGATTAATA  GTTAAGGGACGTGCTTCCCTTAAACACAAGAGAGAAAGAGATTGGCCATGAACATGAAG  AAATGAAGAACCTAGAAGCCATTGTTCAAGAAATAAAACCAACTGCCCTCATAGGAGT  TGCTGCAATTGGTGGTGCATTCTCAGAACAAATTCTCAAAGATATGGCTGCCTTCAAT  GAACGGCCTATTATTTTTGCTTTGAGTAATCCAAGTAGCAAGCAGAATGTTCTGCAG  AGCAGTGCTACAAAATAACCAAGGGACGTGCAATTTTGGCAGTGGCAGTCCCTTTGA  TCCAGTCACTCTTCAAATGGACAGACCCTATATCCTGGCCAAGGCAACAATTCCTAT  GTGTTCCCTGGAGTTGCTTGGTGTGTGGCGTGTGGATTGAGGCAGATCAGAGATA  ATATTTTCTCACTACTGCTGAGGTTATAGCTCAGCAAGTGTGAGATAAACACTTGGA  AGAGGGTCGGCTTTATCCTCCTTTGAATACCATTAGAGATGTTTCTCTGAAAATGCA  GAAAAGATTGTGAAGATGCATACCAAGAAAAGACAGCCACAGTTTATCCTGAACCGC  AAAACAAAGAAGCATTGTCCGCTCCAGATGTATAGTACTGATTATGACCAAGATTCT</p>		

	ACCTGATTGTTATTCTTGGCCTGAAGAGGTGCAGAAAATACAGACCAAAGTTGACCAG TAG		
	ORF Start: at 1		ORF Stop: TAG at 1741
	SEQ ID NO: 210	580 aa	MW at 65129.9kD
NOV14e, 258330149 Protein Sequence	TMGHHHHHHEPEAPRRRTHQRGYLLTRNPHLNKDLAFTLEERQQLNIHGLLPPSFNS QEIQVLRVVKNFEHLNSDFDRYLLMDLQDRNEKLFYRVLTSDIEKFMPIVYTPVGL ACQQYSLVFRKPRGLFITIHDRGHIASVLNAPEDVIKAIVVTDGERILGLDGLGCNG MGIPVGKLALYTACGGMNPQECPLVILDVGTENEELLKDPLYIGLRQRRVRGSEYDDF LDEFMEAVSSSKYGMNCLIQFEDFANVNAFRLNKNYRQYCTFNDDIQGTASVAVAGLL AALRITKNKLSQDQITILFQGAGEAALGIAHLIVMALEKEGLPKEKAIKKIWLVDKGLI VKGRASLTQEKEKFAHEHEEMKNLEAIVQEIKPTALIGVAAIGGAFSEQILKDMAAFN ERP I I FALS NPTS KAEC SAEQCYKITKGRAIFASGSPFDPVTL PNGQTL YPGQGNNSY VFPGVALGVVACGLRQITDNI FLTTAEVIAQQVSDKHLEEGRLYPPLNTIRDVSLKIA EKIVKDAYQEKATATVPEPQNKEAFVRSQMYSTDYDQILPDCYSWPEEVQKIQTQKVDQ		
	SEQ ID NO: 211	1767 bp	
NOV14f, 258330422 DNA Sequence	CACCATCACCACCATCAGAGCCCCGAAGCCCCCGTCGCCGCCACCCATCAGCGCG GCTACCTGCTGACACGGAACCCCTACCTCAACAAGGACTTGGCCTTTACCCTGGAAGA GAGACAGCAATTGAACATTCATGGATTGTTGCCACCTTCCTTCAACAGTCAGGAGATC CAGGTTCTTAGAGTAGTAAAAAATTCGAGCATCTGAACCTTGACTTTGACAGGTATC TTCTCTTAATGGATCTCCAAGATAGAAATGAAAACTCTTTATAGAGTGCTGACATC TGACATTGAGAAATTCATGCCATTGTTTATACTCCCACTGTGGGTCTGGCTTGCCAA CAATATAGTTTGGTGTTCGGAAGCCAAGAGGTCTCTTTATTACTATCCACGATCGAG GGCATATTGCTTCAGTTCTCAATGCATGGCCAGAAGATGTCATCAAGGCCATTGTGGT GACTGATGGAGAGCGTATTCTTGGCTTGGGAGACCTTGGCTGTAATGGAATGGGCATC CCTGTGGGTAAATTTGGCTCTATATACAGCTTGCAGGGGATGAATCCTCAAGAAATGTC TGCTGTGTCATTCTGGATGTGGGAACCGAAAAATGAGGAGTTACTTAAAGATCCACTCTA CATTGGACTACGGCAGAGAAGAGTAAGAGGTTCTGAATATGATGATTTTTTGGACGAA TTCATGGAGGCAGTTTCTTCCAAGTATGGCATGAATTGCCTTATTAGTTTGAAGATT TTGCCAATGTGAATGCATTTCTGCTCCTGAACAAGTATCGAAACAGTATTGCACATT CAATGATGATATTCAAGGAACAGCATCTGTTGCAGTTGCAGGTCTCCTTGCAGCTCTT CGAATAACCAAGAACAACACTGTCTGATCAACAATACTATTCCAAGGAGCTGGGGAGG CTGCCCTAGGGATTGCACACCTGATTGTGATGGCCTTGGAAAAAGAGGTTTACCAA AGAGAAAGCCATCAAAAAGATATGGCTGGTTGATTCAAAAGGATTAATAGTTAAGGGA CGTGCTTCCTTAACACAAGAGAAAGAGAAGTTTGCCCATGAACATGAAGAAATGAAGA ACCTAGAAGCCATTGTTCAAGAAATAAACCAACTGCCCTCATAGGAGTTGCTGCAAT TGGTGGTGCATTCTCAGAACAAATTTCTCAAAGATATGGCTGCCTTCAATGAACGGCCT ATTATTTTGTCTTTGAGTAATCCAAGTAGCAAAGCAGAATGTTCTGCAGAGCAGTGCT ACAAAATAACCAAGGGACGTGCAATTTTGGCAGTGGCAGTCCTTTTGATCCAGTCAC TCTTCCAATGGACAGACCTATATCCTGGCCAAGGCAACAATTCCTATGTGTTCCCT GGAGTTGCTCTTGGTGTGTGGCGTGTGGATTGAGGCAGATCACAGATAATATTTTCC TCACTACTGCTGAGGTTATAGCTCAGCAAGTGTGAGATAAACACTTGAAGAGGGTGC GCTTTATCCTCCTTTGAATACCATTAGAGATGTTTCTCTGAAAATTCAGAAAAGATT GTGAAAGATGCATACCAAGAAAAGACAGCCACAGTTTATCCTGAACCGCAAAACAAAG AAGCATTGTGCCGCTCCAGATGTATAGTACTGATTATGACCAGATTCTACCTGATTG TTATCTTGGCCTGAAGAGGTGCAGAAAATACAGACCAAAGTTGACCAGTAGGCGGCC GCACTCGAGCACCACCACCACCACCAC		
	ORF Start: at 1		ORF Stop: TAG at 1732
	SEQ ID NO: 212	577 aa	MW at 64840.6kD
NOV14f, 258330422 Protein Sequence	HHHHHHHEPEAPRRRTHQRGYLLTRNPHLNKDLAFTLEERQQLNIHGLLPPSFNSQEI QVLRVVKNFELNSDFDRYLLMDLQDRNEKLFYRVLTSDIEKFMPIVYTPVGLACQ QYSLVFRKPRGLFITIHDRGHIASVLNAPEDVIKAIVVTDGERILGLDGLGNGMGI PVGKLALYTACGGMNPQECPLVILDVGTENEELLKDPLYIGLRQRRVRGSEYDDFLDE FMEAVSSSKYGMNCLIQFEDFANVNAFRLNKNYRQYCTFNDDIQGTASVAVAGLLAAL RITKNKLSQDQITILFQGAGEAALGIAHLIVMALEKEGLPKEKAIKKIWLVDKGLIVKG RASLTQEKEKFAHEHEEMKNLEAIVQEIKPTALIGVAAIGGAFSEQILKDMAAFNERP I I FALS NPTS KAEC SAEQCYKITKGRAIFASGSPFDPVTL PNGQTL YPGQGNNSYVFP GVALGVVACGLRQITDNI FLTTAEVIAQQVSDKHLEEGRLYPPLNTIRDVSLKIAEKI		

	VKDAYQEKATATVYPEPONKEAFVRSQMYSTDYDQILPDCYSWPBEVQKIQTQKVDQ		
	SEQ ID NO: 213	1722 bp	
NOV14g, 258330562 DNA Sequence	ACCATGGAGCCCGAAGCCCCCGTCGCCGCCACACCCATCAGCGCGGTACCTGCTGA CACGGAACCCCTCACCTCAACAAGGACTTGGCCTTTACCCTGGAAGAGAGACAGCAATT GAACATTCATGGATTGTTGCCACCTTCCTTCAACAGTCAGGAGATCCAGGTTCTTAGA GTAGTAAAAAATTCGAGCATCTGAACCTCTGACTTTGACAGGTATCTTCTCTTAATGG ATCTCCAAGATAGAAATGAAAACTCTTTATAGAGTGCTGACATCTGACATTGAGAA ATTCATGCCCTATTGTTTATACTCCCACTGTGGGTCTGGCTTGCCAAACAATATAGTTTG GTGTTTCGGAAGCCAAGAGGTCTCTTTATTACTATCCACGATCGAGGGCATATTGCTT CAGTTCTCAATGCATGGCCAGAAGATGTCATCAAGGCCATTGTGGTGACTGATGGAGA GCGTATTCTTGGCTTGGGAGACCTTGGCTGTAATGGAATGGGCATCCCTGTGGGTAAA TTGGCTCTATATACAGCTTGGCGAGGGATGAATCCTCAAGAATGCTGCGCTGCATTTC TGGATGTGGGAACCGAAATGAGGAGTTACTTAAAGATCCACTCTACATTGGACTACG GCAGAGAAGAGTAAGAGGTTCTGAATATGATGATTTTTTGGACGAATTCATGGAGGCA GTTCTCTCAAGTATGGCATGAATTGCCTTATTCAGTTTGAAGATTTTGCCAAATGTGA ATGCATTTCTGCTCTCTGAACAAGTATCGAAACCACTATTGCACATTCAATGATGATAT TCAAGGAACAGCATCTGTTGCAGTTGCAGGTCTCCTTGCAGCTCTTCAATAACCAAG AACAACTGTCTGATCAACAATACTATTCCAAGGAGCTGGAGAGGCTGCCCTAGGGA TTGCACACCTGATTGTGATGGCCTTGGAAGAAAGAGGTTTACCAAAAGAGAAAGCCAT CAAAAGATATGGCTGGTTGATTCAAAGGATTAATAGTTAAGGGACGTGCTTCCTTA ACACAAGAGAAAGAGAGTTTGGCCATGAACATGAAGAAATGAAGAACCTAGAAGCCA TTGTTCAAGAAATAAAACCAACTGCCCTCATAGGAGTTGCTGCAATTGGTGGTGCATT CTCAGAACAAATTCTCAAAGATATGGCTGCCTTCAATGAACGGCCTATTATTTTTGCT TTGAGTAATCCAACCTAGCAAGCAGAATGTTCTGCAGAGCAGTCTACAAAATAACCA AGGGACGTGCAATTTTTGCCAGTGGCAGTCTTTTGATCCAGTCACTCTTCCAATGG ACAGACCCTATATCCTGGCCAAGGCAACAATTCCTATGTGTTCCTGGAGTTGCTCTT GGTGTTGTGGCGTGTGGATTGAGGCAGATCACAGATAATATTTCTCACTACTGCTG AGGTTATAGCTCAGCAAGTGTGAGATAAACACTTGAAGAGGGTTCGGCTTTATCCTCC TTTGAATACCATTAGAGATGTTTCTTGAAAATGTCAGAAAAGATTGTGAAAGATGCA TACCAAGAAAAGACAGCCACAGTTTATCCTGAACCGCAAAACAAAGAAGCATTGTGCC GCTCCAGATGTATAGTACTGATTATGACCAGATCTACCTGATTGTTATCTTGGCC TGAAGAGGTGCAGAAAATACAGACCAAAAGTTGACCAGTAG		
	ORF Start: at 1		ORF Stop: TAG at 1720
	SEQ ID NO: 214	573 aa	MW at 64250.0kD
NOV14g, 258330562 Protein Sequence	TMEPEAPRRRHHQRGYLLTRNPHLNKDLAFTLEERQQLNIHGLLPPSFNSQEIQVLR VVKNFEHLNSDFDRYLLMLDLQRNEKLFYRVLTSDIEKFMPIVYPTVGLACQQYSL VFRKPRGLFITIHDRGHIAVNLNWPEDVIAIVVDGERILGLDLCNGMGIPVKG LALYTACGMNPQECPLPVILDVGTENEELLKDPLYIGLRQRRVRGSEYDDFLDEFMEA VSSKYGMNCLIQFEDFANVNAFRLNKNYRNQYCTFNDDIQGTASVAVAGLLAALRITK NKLSQDQILFQGAGEAALGIAHLIVMALEKEGLPKEKAIKKIWLVDKGLIVKGRASL TQEKKEKFAHEHEEMKNLEAIVQEIKPTALIGVAAIGGAFSEQILKDMAAFNERPIIFA LSNPTSKAECSAEQCYKITKGRAIFASGSPFDPVILPNGQTLYPQGNNSYVFPGVAL GVVACGLRQITDNIFLTAEVIAQQVSDKHLEEGRLYPPLNTIRDVSLKIAEKIVKDA YQEKATATVYPEPONKEAFVRSQMYSTDYDQILPDCYSWPBEVQKIQTQKVDQ		
	SEQ ID NO: 215	1719 bp	
NOV14h, 258330639 DNA Sequence	TGGAGCCCGAAGCCCCCGTCGCCGCCACACCCATCAGCGCGGTACCTGCTGACACG GAACCCCTCACCTCAACAAGGACTTGGCCTTTACCCTGGAAGAGAGACAGCAATTGAAC ATTCATGGATTGTTGCCACCTTCCTTCAACAGTCAGGAGATCCAGGTTCTTAGAGTAG TAAAAAATTCGAGCATCTGAACCTCTGACTTTGACAGGTATCTTCTCTTAATGGATCT CCAAGATAGAAATGAAAACTCTTTATAGAGTGCTGACATCTGACATTGAGAAATTC ATGCCATATTGTTTATACTCCCACTGTGGGTCTGGCTTGCCAAACAATATAGTTTGGTGT TTCGGAAGCCAAGAGGTCTCTTTATTACTATCCACGATCGAGGGCATATTGCTTCAGT TCTCAATGCATGGCCAGAAGATGTCATCAAGGCCATTGTGGTGACTGATCGAGAGCGT ATTCTTGGCTTGGGAGACCTTGGCTGTAATGGAATGGGCATCCCTGTGGGTAAATTGG CTCTATATACAGCTTGGCGAGGGATGAATCCTCAAGAATGCTGCGCTGTATTCTGGA TGTGGGAACCGAAATGAGGAGTTACTTAAAGATCCACTCTACATTGGACTACGGCAG AGAAGAGTAAGAGGTTCTGAATATGATGATTTTTTGGACGAATTCATGGAGGCAGTTT		

	CTTCCAAGTATGGCATGAATTGCCTTATTTCAGTTTGAAGATTTTGCCAAATGTGAATGC ATTTTCGTCTCCTGAACAAGTATCGAAACCAGTATTGCACATTCAATGATGATATTCAA GGAACAGCATCTGTTGCAGTTGCAGGTCTCCTTGCAGCTCTTGAATAACCAAGAACA AACTGTCTGATCAAACAATACTATTCCAAGGAGCTGGGGAGGCTGCCCTAGGGATTGC ACACCTGATTGTGATGGCCTTGAAAAAGAAGGTTTACCAAAAGAGAAAGCCATCAAA AAGATATGGCTGGTTGATTCAAAAGGATTAATAGTTAAGGGACGTGCTTCCTTAACAC AAGAGAAAGAGAAGTTTGCCCATGAACATGAAGAAATGAAGAACCTAGAAGCCATTGT TCAAGAAATAAAACCAACTGCCCTCATAGGAGTTGCTGCAATTGGTGGTGCATTCTCA GAACAAATTCTCAAAGATATGGCTGCCTTCAATGAACGGCCTATTATTTTGTCTTGA GTAATCCAAC TAGCAAAGCAGAATGTTCTGCAGAGCAGTGCTACAAAATAACCAAGGG ACGTGCAATTTTGGCAGTGGCAGTCCTTTTGATCCAGTCACTTTCCTCAATGGACAG ACCCTATATCCTGGCCAAGGCAACAATTCCTATGTGTTCCCTGGAGTTGCTCTTGGTG TTGTGGCGTGTGGATTGAGGCAGATCACAGATAATATTTTCTCACTACTGCTGAGGT TATAGCTCAGCAAGTGTGAGATAAACAATTGGAAGAGGGTCGGCTTTATCCTCCTTTG AATACCATTAGAGATGTTTCTCTGAAAATTGCAGAAAAGATTGTGAAAGATGCATACC AAGAAAAGACAGCCACAGTTTATCCTGAACCGCAAAACAAAGAAGCAATTGTCCGCTC CCAGATGTATAGTACTGATTATGACCAGATTCTACCTGATTGTTATTCTTGGCCTGAA GAGGTGCAGAAAATACAGACCAAGTTGACCAGTAGG		
	ORF Start: at 3		ORF Stop: TAG at 1716
	SEQ ID NO: 216	571 aa	MW at 64017.7kD
NOV14h, 258330639 Protein Sequence	EPEAPRRRHTHQRGYLLTRNPHLNKDLAFTLEERQQLNIHGLLPSPFNSQEIQVLRVV KNFEHLNSDFDRYLLMDLQDRNEKLFYRVLTSIEKFMPIVYPTVGLACQQYSLVF RKPRGLFTITHDRGHIAVLNAPEDVIKATVVDGERILGLDGLGCMGMI PVGKLA LYTACGGMNPQECLPVILDVGTENEELLKDPYIGLRQRRVRGSEYDDFLDEFMEAVS SKYGMNCLIQFEDFANVNAFRLNKYRNQYCTFNDDIQGTASVAVAGLLAALRITKNK LSDQITILFQGAGEAALGIAHLIVMALEKEGLPKEKAIKKIWLVDKGLIVKGRASLTQ EKEKFAHEHEEMKNLEAIVQEIKPTALIGVAAIGGAFSEQILKDMAAFNERPIIFALS NPTSKAECSAEQCYKITKGRAIFASGSPFDVTL PNGQTLYPGQGNNSYVFPGVALGV VACGLRQITDNI FLTTAEVIAQQVSDKHLEEGRLYPPLNTIRDVSLKIAEKIVKDAYQ EKTATVYEPQNKAEFVRSQMYSTDYDQILPDCYSWPBEEVQKIQTQKVDQ		
	SEQ ID NO: 217	1732 bp	
NOV14i, 259357792 DNA Sequence	ACCATGGAGCCCGAAGCCCCCGTCGCCGCCACCCATCAGCGCGGTACCTGCTGA CACGGAACCCCTCACCTCAACAAGGACTTGGCCTTTACCCTGGAAGAGAGACAGCAATT GAACATTCATGGATTGTTGCCACCTTCCTTCAACAGTCAAGGATCCAGGTTCTTAGA GTAGTAAAAAATTTCCAGCATCTGAACTCTGACTTTGACAGGTATCTTCTCTTAATGG ATCTCCAAGATAGAAATGAAAACTCTTTATAGAGTGCTGACATCTGACATTGAGAA ATTCATGCCATTGTTTATATCTCCCACTGTGGGTCTGGCTTGGCAACAATATAGTTG GTGTTTCGGAAGCCAAGAGGTCTCTTTATTACTATCCACGATCGAGGGCATATTGCTT CAGTTCTCAATGCATGGCCAGAAGATGTCAAGGCCATTGTGGTGACTGATGGAGA GCGTATCTTGGCTTGGGAGACCTTGGCTGTAATGGAATGGGCATCCCTGTGGGTAAA TTGGCTCTATATACAGCTTGGCGAGGGATGAATCCTCAAGAAATGTCTGCCTGTCAATC TGGATGTGGGAACCGAAAATGAGGAGTTACTTAAAGATCCACTACATTGGACTACG GCAGAGAAGAGTAAGAGGTTCTGAATATGATGATTTTTTGGACGAATTCATGGAGGCA GTTTCTTCCAAGTATGGCATGAATTGCCCTTATTTCAGTTTGAAGATTTTGCCAATGTGA ATGCATTTTCGTCTCCTGAACAAGTATCGAAACCAGTATTGCACATTCAATGATGATAT TCAAGGAACAGCATCTGTTGCAGTTGCAGGTCTCCTTGAGCTCTTCAATAACCAAG AACAACTGTCTGATCAACAATACTATTCCAAGGAGCTGGGGAGGCTGCCCTAGGGA TTGCACACCTGATTGTGATGGCCTTGAAAAAGAAGGTTTACCAAAAGAGAAAGCCAT CAAAAAGATATGGCTGGTTGATTCAAAAGGATTAATAGTTAAGGGACGTGCTTCCTTA ACACAAGAGAAAGAGAAGTTTGCCCATGAACATGAAGAAATGAAGAACCTAGAAGCCA TTGTTCAAGAAATAAAACCAACTGCCCTCATAGGAGTTGCTGCAATTGGTGGTGCATT CTCAGAACAAATCTCAAAGATATGGCTGCCTTCAATGAACGGCCTATTATTTTGGCT TTGAGTAATCCAAC TAGCAAAGCAGAATGTTCTGCAGAGCAGTGCTACAAAATAACCA AGGGACGTGCAATTTTGGCAGTGGCAGTCCTTTTGATCCAGTCACTTCTTCAATGG ACAGACCCTATATCCTGGCCAAGGCAACAATTCCTATGTGTTCCCTGGAGTTGCTCTT GGTGTGTGGCGTGTGGATTGAGGCAGATCACAGATAATATTTTCTCACTACTGCTG AGGTTATAGCTCAGCAAGTGTGAGATAAACAATTGGAAGAGGGTCGGCTTTATCCTCC TTTGAATACCATTAGAGATGTTTCTCTGAAAATTGCAGAAAAGATTGTGAAAGATGCA		

	TACCAAGAAAAGACAGCCACAGTTTATCCTGAACCGCAAAACAAAGAAGCATTGTCC GCTCCAGATGTATAGTACTGATTATGACCAGATTCTACCTGATTGTTATTCTTGGCC TGAAGAGGTGCAGAAAATACAGACCAAAGTTGACCAGTAGGCGGCCGCTT		
	ORF Start: at 1		ORF Stop: TAG at 1720
	SEQ ID NO: 218	573 aa	MW at 64250.0kD
NOV14i, 259357792 Protein Sequence	TMEPEAPRRRHTHQRGYLLTRNPHLNKDLAFTLEERQQLNIHGLLPPSFNSQEIQVLR VVKNFHLSNDFDRYLLMDLQDRNEKLFYRVLTSDIEKFMPIVYTPTVGLACQQYSL VFRKPRGLFITIHDRGHASVLNAPEDVIKAIIVTDGERILGLDGLCGNGMGI PVGK LALYTACGGMNPQECLPVILDVGTENEELLKDPLYIGLRQRRVRGSEYDDFLDEFMEA VSSKYGMNCLIQFEDFANVNAFRLLNKYRNQYCTFNDDIQGTASVAVAGLLAALRITK NKLSDQITILFQAGEAALGIAHLIVMALEKEGLPKEKAIKKIWLVD SKGLIVKGRASL TQEKEKFAHEHEEMKNLEAIVQEIKPTALIGVAAIGGAFSEQILKDMAAFNERPIIFA LSNPTSKAECSAEQCYKITKGRAIFASGSPFDPVTL PNGQTLYPGQGNNSYVFPGVAL GVVACGLRQITDNI FLTTAEVIAQQVSDKHLEEGRLYPPLNTRDVS LKIAEKIVKDA YQEKATATVYPEPQNKEAFVRSQMYSTDYDQILPDCYSWPPEEVOKIQTKVDQ		
	SEQ ID NO: 219	1838 bp	
NOV14j, CG140316-02 DNA Sequence	CCGGCGCCAGCCATGGAGCCCGAAGCCCCCGTCGCGCCACACCCATCAGCGCGGCT ACCTGCTGACACGGAACCTCACCTCAACAAGGACTTGGCCTTTACCCTGGAAGAGAG ACAGCAATTGAACATTCATGGATTGTGCCACCTTCTTCAACAGTCAGGAGATCCAG GTTCTTAGAGTAGTAAAAAATTCGAGCATCTGAACCTGACTTTGACAGGTATCTTC TCTTAATGGATCTCCAAGATAGAAAATGAAAACTCTTTTATAGAGTGTGCATCTGA CATTGAGAAATTCATGCCTATTGTTTATACTCCCACTGTGGGTCTGGCTTGCCAACAA TATAGTTTGGTGTTCGGAAGCCAAGAGGTCTCTTTATTACTATCCACGATCGAGGGC ATATTGCTTCAGTTCTCAATGCATGCCAGAGATGTGCATCAAGGCCATTGTGGTGAC TGATGAGAGCGTATTCTTGGCTTGGGAGACCTTGGCTGTAATGGAATGGGCATCCCT GTGGGTAAATTGGCTCTATATACAGCTTGGGAGGGATGAATCCTCAAGAATGTCTGC CTGTCAATCTGGATGTGGGAACCGAAAAATGAGGAGTTACTTAAAGATCCACTTACAT TGGACTACGGCAGAGAAGAGTAAGAGGTTCTGAATATGATGATTTTTTGGACGAATTC ATGGAGGCAGTTTCTTCCAAGTATGGCATGAATGTCCTTATTCAGTTTGAAGATTTTG CCAATGTGAATGCATTTCTGCTCCTGAACAAGTATCGAAACCAAGTATTCACATTCAA TGATGATATTCAAGGAACAGCATCTGTTGCAGTGCAGGTCTCCTTGCAGCTCTTCGA ATAACCAAGAACAACTGTCTGATCAAAACAATACTATTCCAAGGAGCTGGAGAGGCTG CCCTAGGGATTGCACACCTGATTGTGATGGCCTTGAAAAAGAAGGTTTACCAAGGA GAAAGCCATCAAAAAGATATGGCTGGTTGATTCAAAAGGATTAATAGTTAAGGGACGT GCTTCCTTAACACAAGAGAAAGAGAAGTTTGCCCATGAACATGAAGAAATGAAGAAC TAGAAGCCATTGTTCAAGAAATAAAACCAACTGCCCTCATAGGAGTTGCTGCAATTGG TGGTGCATTCTCAGAACAAATCTCAAAGATATGGCTGCCTTCAATGACGCGCTATT ATTTTTGCTTTGAGTAATCCAAGTAGCAAGCAGAATGTTCTGCAGAGCAGTGCTACA AAATAACCAAGGGACGTGCAATTTTGCCAGTGGCAGTCTTTTGATCCAGTCACTCT CCCAGATGGACGGACTCTGTTTCTGGCCAAGGCAACAATTCCTACGTGTTCCCTGGA GTTGCTCTTGGGGTGGTGGCTGCGGACTGAGACACATCGATGATAAGGTCTTCCTCA CCACTGCTGAGGTATATCTCAGCAAGTGTGAGATAAACACCTGCAAGAAGGCCGCT CTATCCTCCTTTGAATACCATTTCGAGACGTTTCGTTGAAAATTGCAGTAAAGATTGTG CAAGATGCATACAAAGAAAAGATGGCCACTGTTTATCCTGAACCCCAAAACAAAGAAG AATTTGTCTCCTCCAGATGTACAGCACTAATATGACCAGATCCTACCTGATTGTTA TCCGTGGCCTGCAGAAGTCCAGAAAATACAGACCAAAGTCAACCAGTAACGCAACAGC TAGGATTTTTAACTTTATTAGTAAAATCTTGAAGTTTTTCATGATCTTTAAGGGTCAG ATCTTTTATGATGATTATGATGATGCTTAGAATAAGGTGC		
	ORF Start: ATG at 13		ORF Stop: TAA at 1729
	SEQ ID NO: 220	572 aa	MW at 64139.1kD
NOV14j, CG140316-02 Protein Sequence	MEPEAPRRRHTHQRGYLLTRNPHLNKDLAFTLEERQQLNIHGLLPPSFNSQEIQVLRV VKNFEHLNSDFDRYLLMDLQDRNEKLFYRVLTSDIEKFMPIVYTPTVGLACQQYSLV FRKPRGLFITIHDRGHASVLNAPEDVIKAIIVTDGERILGLDGLCGNGMGI PVGKL ALYTACGGMNPQECLPVILDVGTENEELLKDPLYIGLRQRRVRGSEYDDFLDEFMEA SSKYGMNCLIQFEDFANVNAFRLLNKYRNQYCTFNDDIQGTASVAVAGLLAALRITKN KLSDQITILFQAGEAALGIAHLIVMALEKEGLPKEKAIKKIWLVD SKGLIVKGRASLT QEKEKFAHEHEEMKNLEAIVQEIKPTALIGVAAIGGAFSEQILKDMAAFNERPIIFAL		

	SNPTSKAECSAEQCYKITKGRAIFASGSPFDPVTLPDGRTLFPQGNNNSYVFPGVALG VVACGLRHIDDKVFLTTAEVISQQVSDKHLQEGRLYPPLNTIRDVSLKIAVKIVQDAY KEKMATVYPEQNKEEFVSSQMYSTNYDQILPDCYPWPAEVQKIQTQKVNQ		
	SEQ ID NO: 221	1750 bp	
NOV14k, CG140316-03 DNA Sequence	CGCGGATCCATGGAGCCCGAAGCCCCCGTCGCCGCCACACCCATCAGCGCGGCTACC TGCTGACACGGAACCCCTCACCTCAACAAGGACTTGGCCTTTACCCCTGGAAGAGAGACA GCAATTGAACATTCATGGATTGTTGCCACCTTCCTTCAACAGTCAGGAGATCCAGGTT CTTAGAGTAGTAAAAAATTTGAGCATCTGAACTCTGACTTTGACAGGTATCTTCTCT TAATGGATCTCCAAGATAGAAATGAAAACTCTTTTATAGAGTGCTGACATCTGACAT TGAGAAATTCATGCCTATTGTTTATACTCCCACTGTGGGTCTGGCTTGCCAACAATAT AGTTTGGTGTTCGGAAGCCCAAGAGGTCTCTTTATTACTATCCACGATCGAGGGCATA TTGCTTCAGTTCTCAATGCATGGCCAGAAGATGTCATCAAGGCCATTGTGGTGACTGA TGGAGAGCGTATCTTGGCTTGGGAGACCTTGGCTGTAATGGAATGGGCATCCCTGTG GGTAAATTGGCTCTATATACAGCTTGGCGAGGGATGAATCCTCAAGAATGTCTGCCTG TCATTCTGGATGTGGGAACCGAAAAATGAGGAGTTACTTAAAGATCCACTCTACATTGG ACTACGGCAGAGAAGAGTAAGAGGTTCTGAATATGATGATTTTTTGGACGAATTCATG GAGGCAGTTTCTTCCAAGTATGGCATGAATTGCCTTATTTCAGTTTGAAGATTTTGCCA ATGTGAATGCATTTCGTCTCCTGAACAAGTATCGAAACAGTATTGCACATTCAATGA TGATATTCAAGGAACAGCATCTGTTGCAGTTGCAGGTCTCCTTGCAGCTCTTCGAATA ACCAAGAACAACCTGTCTGATCAAACAATACTATTCCAAGGAGCTGGGAGGCTGCC TAGGGATTGCACACCTGATGTGTGATGGCCTTGAAAAAGAAGGTTTACCAAAAGAGAA AGCCATCAAAAAGATATGGCTGGTTGATTCAAAAGGATTAATAGTTAAGGGACGTGCT TCCTTAACACAAGAGAAAGAGAAGTTTGCCCATGAACATGAAGAAATGAAGAACCTAG AAGCCATTGTTCAAGAAATAAAACCAACTGCCCTCATAGGAGTTGCTGCAATTGGTGG TGCATTCTCAGAACAAATTCCTAAAGATATGGCTGCCTTCAATGAACGGCCTATTATT TTTGCTTTGAGTAATCCAACCTAGCAAAGCAGAATGTTCTGCAGAGCAGTGCTACAAAA TAACCAAGGGACGTGCAATTTTGCCAGTGGCAGTCCTTTTGATCCAGTCACCTTCC AAATGGACAGACCCTATATCCTGGCCAAGGCAACAATTCCTATGTGTTCCCTGGAGTT GCTCTTGGTGTGTGGCGTGTGGATTGAGGCAGATCAGAGATAATATTTTCCCTACTA CTGCTGAGGTTATAGCTCAGCAAGTGTGAGATAAACTTGAAGAGGGTCCGCTTTA TCCTCCTTTGAATACCATTAGAGATGTTTCTCTGAAAATTGCAGAAAAGATTGTGAAA GATGCATACCAAGAAAAGACAGCCACAGTTTATCCTGAACCGCAAAACAAAGAGCAT TTGTCCGCTCCCAGATGTATAGTACTGATTATGACCAGATTCTACCTGATTGTTATTC TTGGCCTGAAGAGGTGCAGAAAATACAGACCAAAAGTTGACCAGTAGGGTGGCGGCCG TTTTTCCTT		
	ORF Start: ATG at 10		ORF Stop: TAG at 1726
	SEQ ID NO: 222	572 aa	MW at 64148.9kD
NOV14k, CG140316-03 Protein Sequence	MEPEAPRRRHHQRGYLLTRNPHLNKDLAFTLEERQQLNIHGLLPPSFNSQETQVLRV VKNFEHLNSDFDRYLLMLDQRNEKLFYRVLTSDIEKFMPIVYTPTVGLACQQYSLV FRKPRGLFITIHDRGHIAVLNAPEDVIKAIVVTDGERILGLDLGCNGMGI PVGKL ALYTACGMNPQECLPVILDVGTENEELLKDPLYIGLRQRRVRGSEYDDFLDEFMEAV SSKYGMNCLIQFEDFANVNAFRLLNRYRNYCTFNDDIQGTASVAVAGLLAALRITKN KLSQDQILFQAGAEALGIAHLIVMALEKEGLPKKAIKKIWLVDKGLIVKGRASLT QEKEKFAHEHEEMKNLEAIVQEI KPTALIGVAAIGGAFSEQILKDMAAFNERPIIFAL SNPTSKAECSAEQCYKITKGRAIFASGSPFDPVTL PNGQTLYPQGNNNSYVFPGVALG VVACGLRQITDNIPLTTAEVIAQQVSDKHLQEGRLYPPLNTIRDVSLKIAEKIVKDAY QEKATATVYPEQNKEAFVRSQMYSTDYDQILPDCYSWPBEVQKIQTQKVDQ		
	SEQ ID NO: 223	1767 bp	
NOV14l, CG140316-04 DNA Sequence	CACCATCACCACCATCAGAGCCCGAAGCCCCCGTCGCCGCCACACCCATCAGCGG GCTACCTGCTGACACGGAACCCCTCACCTCAACAAGGACTTGGCCTTTACCCCTGGAAGA GAGACAGCAATTGAACATTCATGGATTGTTGCCACCTTCCTTCAACAGTCAGGAGATC CAGGTTCTTAGAGTAGTAAAAAATTTGAGCATCTGAACTCTGACTTTGACAGGTATC TTCTCTTAATGGATCTCCAAGATAGAAATGAAAACTCTTTTATAGAGTGCTGACATC TGACATTGAGAAATTCATGCCTATTGTTTATACTCCCACTGTGGGTCTGGCTTGCCAA CAATATAGTTTGGTGTTCGGAAGCCCAAGAGGTCTCTTTATTACTATCCACGATCGAG GGCATATTGCTTCAGTTCTCAATGCATGGCCAGAAGATGTCATCAAGGCCATTGTGGT GACTGATGGAGAGCGTATTCTTGGCTTGGGAGACCTTGGCTGTAATGGAATGGGCATC		

	CCTGTGGGTAAATTGGCTCTATATACAGCTTGCGGAGGGATGAATCCTCAAGAATGTC TGCCTGTCAATCTGGATGTGGGAACCGAAAATGAGGAGTTACTTAAAGATCCACTCTA CATTTGACTACGGCAGAGAAGAGTAAGAGGTTCTGAATATGATGATTTTTTGGACGAA TTCATGGAGGCAGTTTCTTCCAAGTATGGCATGAATTGCCTTATTCAGTTTGAAGATT TTGCCAATGTGAATGCATTTCTGCTCCTGAACAAGTATCGAAACCAGTATTGCACATT CAATGATGATATTCAAGGAACAGCATCTGTTGCAGTTGCAGGTCCTCTTCAGACTCTT CGAATAACCAAGAACAACACTGTCTGATCAAACAATACTATTCCAAGGAGCTGGGGAGG CTGCCCTAGGGATTGCACACCTGATTGTGATGGCCTTGGAAGAAAGAGTTTACCAA AGAGAAAGCCATCAAAAAGATATGGCTGGTTGATTCAAAAGGATTAATAGTTAAGGGA CGTGCTTCCTTAACACAAGAGAAAGAGAAGTTTGGCCATGAACATGAAGAAATGAAGA ACCTAGAAGCCATTGTTCAAGAAATAAAACCACTGCCCTCATAGGAGTTGCTGCAAT TGGTGGTGCATTCTCAGAACAATTTCTCAAAGATATGGCTGCCCTTCAATGAACGGCCT ATTATTTTGTCTTTGAGTAATCCAACCTAGCAAAGCAGAATGTTCTGCAGAGCAGTGCT ACAAAATAACCAAGGGACGTGCAATTTTGGCAGTGGCAGTCCTTTTGATCCAGTCAC TCTTCCAATGGACAGACCTTATATCCTGGCCAAGGCAACAATTCCTATGTGTTCCCT GGAGTTGCTCTTGGTGTGTTGGCGTGTGGATTGAGGCAGATCACAGATAATATTTTCC TCACTACTGCTGAGGTTATAGCTCAGCAAGTGTGAGATAAACAAGTGAAGAGGGTCTG GCTTTATCCTCCTTTGAATACCATTAGAGATGTTCTCTGAAAATTGCAGAAAAGATT GTGAAAGATGCATACCAAGAAAAGACAGCCACAGTTTATCCTGAACCGCAAAACAAAG AAGCATTTGTCCGCTCCAGATGTATAGTACTGATTATGACCAGATTCTACCTGATTG TTATTCTTGGCCTGAAGAGGTGCAGAAAATACAGACCAAGTTGACCAGTAGGCGGCC GCACTCGAGCACCACCACCACCAC		
	ORF Start: at 1		ORF Stop: TAG at 1732
	SEQ ID NO: 224	577 aa	MW at 64840.6kD
NOV14l, CG140316-04 Protein Sequence	HHHHHHEPEAPRRRTHQRGYLLTRNPHLNKDIAFTLEERQQLNIHGLLPSPFNSQEI QVLRVVKNFPEHLNSDFDRYLLMLDQDRNEKLFYRVLTSDIEKFMPIVYPTVGLACQ QYSLVFRKPRGLFITIHDRGHIAVNLAWPEDVIKAIIVTDTGERILGLDGLGNGMGI PVGKLALYTACGMNPQECPLVILDVGTENEELKDPPLYIGLRQRRVRGSEYDDFLDE FMEAVSSKYGMNCLIQFEDFANVNAFRLLNKYRNQYCTFNDDIQGTASVAVAGLLAAL RITKNKLSQDITLFGAGEAALGIAHLIVMALEKEGLPKEKAIKKIWLVDKGLIVKG RASLTQEKEKFAHEHEEMKNLEATVQEI KPTALIGVAAIGGAFSEQILKDMAAFNERP IIFALSNPTSKAECSAEQCYKITKGRAIFASGSPFDPVTLPLNGQTLPGQGNNSYVFP GVALGVVACGLRQITDNI FLTTAEVIAQQVSDKHLEEGRLYPLNTIRDVSLKIAEKI VKDAYQEKTATVYPEPQNKEAFVRSQMYSYTDYDQILPDCYSWPEEVQKIQTQKVDQ		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 14B.

Table 14B. Comparison of NOV14a against NOV14b through NOV14l.		
Protein Sequence	NOV14a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV14b	1..572	572/572 (100%)
	1..572	572/572 (100%)
NOV14c	1..572	572/572 (100%)
	4..575	572/572 (100%)
NOV14d	3..572	570/570 (100%)
	1..570	570/570 (100%)
NOV14e	2..572	571/571 (100%)
	10..580	571/571 (100%)
NOV14f	2..572	571/571 (100%)
	7..577	571/571 (100%)

NOV14g	1..572 2..573	572/572 (100%) 572/572 (100%)
NOV14h	2..572 1..571	571/571 (100%) 571/571 (100%)
NOV14i	1..572 2..573	572/572 (100%) 572/572 (100%)
NOV14j	1..572 1..572	553/572 (96%) 563/572 (97%)
NOV14k	1..572 1..572	572/572 (100%) 572/572 (100%)
NOV14l	2..572 7..577	571/571 (100%) 571/571 (100%)

Further analysis of the NOV14a protein yielded the following properties shown in Table 14C.

Table 14C. Protein Sequence Properties NOV14a	
PSort analysis:	0.7000 probability located in nucleus; 0.3000 probability located in microbody (peroxisome); 0.1771 probability located in lysosome (lumen); 0.1000 probability located in mitochondrial matrix space
SignalP analysis:	No Known Signal Sequence Predicted

5 A search of the NOV14a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 14D.

Table 14D. Geneseq Results for NOV14a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV14a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAR52605	Human cytoplasmic NADP+-dependent malate enzyme ME1 - Homo sapiens, 572 aa. [EP595241-A, 04-MAY-1994]	1..572 1..572	572/572 (100%) 572/572 (100%)	0.0
AAM40228	Human polypeptide SEQ ID NO 3373 - Homo sapiens, 604 aa. [WO200153312-A1, 26-JUL-2001]	13..568 48..603	404/556 (72%) 485/556 (86%)	0.0
AAU33270	Novel human secreted protein #3761 - Homo sapiens, 621 aa.	13..568 58..620	380/563 (67%) 464/563 (81%)	0.0

	[WO200179449-A2, 25-OCT-2001]			
AAM42014	Human polypeptide SEQ ID NO 6945 - Homo sapiens, 624 aa. [WO200153312-A1, 26-JUL-2001]	13..568 58..623	376/566 (66%) 458/566 (80%)	0.0
ABG21889	Novel human diagnostic protein #21880 - Homo sapiens, 625 aa. [WO200175067-A2, 11-OCT-2001]	13..568 58..624	372/567 (65%) 455/567 (79%)	0.0

In a BLAST search of public sequence databases, the NOV14a protein was found to have homology to the proteins shown in the BLASTP data in Table 14E.

Table 14E. Public BLASTP Results for NOV14a				
Protein Accession Number	Protein/Organism/Length	NOV14a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
P48163	NADP-dependent malic enzyme (EC 1.1.1.40) (NADP-ME) (Malic enzyme 1) - Homo sapiens (Human), 572 aa.	1..572 1..572	572/572 (100%) 572/572 (100%)	0.0
Q16797	NADP-dependent malic enzyme (EC 1.1.1.40) - Homo sapiens (Human), 572 aa.	1..572 1..572	553/572 (96%) 563/572 (97%)	0.0
JC4160	malate dehydrogenase (oxaloacetate-decarboxylating) (NADP+) (EC 1.1.1.40) - human, 572 aa.	1..572 1..572	552/572 (96%) 562/572 (97%)	0.0
P13697	NADP-dependent malic enzyme (EC 1.1.1.40) (NADP-ME) (Malic enzyme 1) - Rattus norvegicus (Rat), 572 aa.	1..572 1..572	517/572 (90%) 549/572 (95%)	0.0
Q921S3	Malic enzyme, supernatant - Mus musculus (Mouse), 572 aa.	1..572 1..572	516/572 (90%) 545/572 (95%)	0.0

PFam analysis predicts that the NOV14a protein contains the domains shown in the Table 14F.

Table 14F. Domain Analysis of NOV14a			
Pfam Domain	NOV14a Match Region	Identities/ Similarities	Expect Value

		for the Matched Region	
Paramyx_ncap	278..314	14/37 (38%) 24/37 (65%)	0.77
malic	15..553	356/580 (61%) 515/580 (89%)	0

**Example 15.**

The NOV15 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 15A.

Table 15A. NOV15 Sequence Analysis			
	SEQ ID NO: 225	4427 bp	
NOV15a, CG142427-01 DNA Sequence	GGCACGAGGCCGGGACAAAAGCCGGATCCCGGGAAGCTACCGGCTGCTGGGGTGCTCC GGATTTTGC GG GTTCGT CGGG CCTGTGGAAGAAGCGCCGCGCAGGACTTCGGCAGA GGTAGAGCAGGTCTCTCTGCAGCCATGT CGGCCAAGGCAATTCAGAGCAGACGGGCA AAGAACTCCTTTACAAGTTCATCTGTACCACCTCAGCCATCCAGAATCGGTTCAAGTA TGCTCGGGTCACTCCTGACACAGACTGGGCCCGCTTGCTGCAGGACCACCCCTGGCTG CTCAGCCAGAACTTGGTAGTCAAGCCAGACCAGCTGATCAAACGTCGTGAAAACTTG GTCTCGTTGGGGTCAACCTCACTCTGGATGGGGTCAAGTCCTGGCTGAAGCCACGGCT GGGACAGGAAGCCACAGTTGGCAAGGCCACAGGCTTCCTCAAGAACCTTCTGATCGAG CCCTTCGTCCCCACAGTCAGGCTGAGGAGTTCTATGTCTGCATCTATGCCACCCGAG AAGGGGACTACGTCTGTTCACCCAGAGGGGGGTGTGGACGTGGGTGATGTGGACGC CAAGGCCCAGAAGCTGCTTGTGGCGTGGATGAGAACTGAATCCTGAGGACATCAAA AAACACCTGTTGGTCCACGCCCTGAAGACAAGAAAGAAATCTGGCCAGTTTATCT CCGGCCTCTTCAATTTCTACGAGGACTTGTACTTCACTTACCTCGAGATCAATCCCT TGTAAGTACCAAAGATGGAGTCTATGTCTTGACTTGGCGGCCAAGGTGGACGCCACT GCCGACTACATCTGCAAAGTGAAGTGGGGTGACATCGAGTTCCCTCCCCCTTCGGGC GGGAGGCATATCCAGAGGAAGCCTACATTGACAGCCTCGATGCCAAAGTGGGGCAAG CCTGAAGCTGACCTTGCTGAACCCCAAAGGAGGATCTGGACCATGGTGGCCGGGGGT GGCGCCTCTGTCTGTACAGCGATACCATCTGTGATCTAGGGGTGTCAACGAGCTGG CAAATATGGGGAGTACTCAGGCGCCCCAGCGAGCAGCAGACCTATGACTATGCCAA GACTATCCTCTCCCTCATGACCCGAGAGAAGCACCCAGATGGCAAGATCCTCATCATT GGAGGCAGCATCGCAAACCTTCAACACGTGGCTGCCACGTTCAAGGGCATCGTGAGAG CAATTCGAGATTACAGGGCCCCCTGAAGGAGCACGAAGTCACAATCTTTGTCCGAAG AGGTGGCCCCAATATCAGGAGGGCTTACGGGTGATGGGAGAAGTCGGGAAGACCACT GGGATCCCCATCCATGTCTTTGGCACAGAGACTCACATGACGGCCATTTGTGGGCATGG CCCTGGGCCACCGGCCCATCCCCAACCAGCCACCCACAGCGGCCACACTGCAAACCTT CCTCCTCAACGCCAGCGGGAGCACATCGACGCCAGCCCCAGCAGGACAGCATCTTTT TCTGAGTCCAGGGCCGATGAGGTGGCGCCTGCAAAGAAGGCCAAGCCTGCCATGCCAC AAGATTCAGTCCCAAGTCCAAGATCCCTGCAAGGAAAGAGCACCACCTCTTCAGCCG CCACACCAAGGCCATTGTGTGGGGCATGCAGACCCGGGCGGTGCAAGGCATGTGGAC TTTGAATATGTCTGCTCCCAGACGAGCCCTCAGTGGCTGCCATGGTCTACCTTTCA CTGGGGACCACAAGCAGAAGTTTACTGGGGGCACAAAGAGATCCTGATCCCTGTCTT CAAGAACATGGCTGATGCCATGAGGAAGCATCCGGAGGTAGATGTGTCATCAACTTT GCCTCTCTCCGCTCTGCCTATGACAGCACCATGGAGACCATGAACATGCCCAGATCC GGACCATCGCCATCATAGCTGAAGGCATCCCTGAGGCCCTCACGAGAAGCTGATCAA GAAGGCGGACCAGAAGGGAGTGACCATCATCGGACCTGCCACTGTTGGAGGCATCAAG CCTGGGTGCTTTAAGATTGGCAACACAGGTGGGATGCTGGACAACATCCTGGCCTCCA AACTGTACCGCCAGGCAGCGTGGCCTATGTCTCACGTTCCGGAGGCATGTCCAACGA GCTCAACAATATCATCTCTCGGACCACGGATGGCGTCTATGAGGGCGTGGCCATTTGGT GGGGACAGGTACCCGGGCTCCACATTCATGGATCATGTGTTACGCTATCAGGACACTC CAGGAGTCAAAATGATTGTGTTCTTGGAGAGATTGGGGGCATGAGGAATATAAGAT TTGCCGGGGCATCAAGGAGGGCCGCTCACTAAGCCCATCGTCTGCTGGTGCATCGGG ACGTGTGCCACCATGTTCTCTCTGAGGTCCAGTTTGGCCATGCTGGAGCTTGTGCCA		

	ACCAGGCTTCTGAACTGCAGTAGCCAAGAACCAGGCTTGAAGGAAGCAGGAGTGT TGTGCCCCGAGCTTTGATGAGCTTGGAGAGATCATCCAGTCTGTATACGAAGATCTC GTGGCCAATGGAGTCATTGTACCTGCCAGGAGGTGCCGCCCCAACCGTGCCCATGG ACTACTCTCTGGGCCAGGGAGCTTGGTTTGATCCGCAAACTGCCTCGTTCATGACCAG CATCTGCGATGAGCGAGGACAGGAGCTCATCTACGCGGGCATGCCCATCACTGAGGTC TTCAAGGAAGAGATGGGCATTGGCGGGGTCTCGGCCTCTCTGGTTCCAGAAAAGGT TGCCTAAGTACTCTTGCCAGTTCATTGAGATGTGTCTGATGGTGACAGCTGATCAGCG GCCAGCCGTCTCTGGAGCCCAACACCATCATTTGTGCGCGAGCTGGGAAAGACCTG GTCTCCAGCCTCACTCGGGGCTGCTCACCATCGGGGATCGGTTTGGGGGTGCCTTGG ATGCAGCAGCCAAGATGTTCACTAAAGCCTTTGACAGTGGCATTATCCCCATGGAGTT TGTGAACAAGATGAAGAAGGAAGGAAGCTGATCATGGGCATTGGTCACCGAGTGAAG TCGATAAACCAACCCAGACATGCGAGTGCAGATCCTCAAAGATTACGTCAGGCAGCACT TCCCTGCCACTCCTCTGCTCGATTATGCACTGGAAGTAGAGAAGATTACCACCTCGAA GAAGCCAAATCTTATCCTGAATGTAGATGGTCTCATCGAGTCGCATTTGTAGACATG CTTAGAAACTGTGGGTCTTTACTCGGGAGGAAGCTGATGAATATATTGACATTGGAG CCCTCAATGGCATCTTTGTGCTGGGAAGGAGTATGGGGTTCATTGGACACTATCTTGA TCAGAAGAGGCTGAAGCAGGGGCTGTATCGTCATCCGTGGGATGATTTTCATATGTT CTTCCGGAACACATGAGCATGTAAACAGAGCCAGGAACCCCTACTGCAGTAAACTGAAGA CAAGATCTCTTCCCCAAGAAAAAGTGTACAGACAGCTGGCAGTGGAGCCTGCTTTAT TTAGCAGGGGCTTGAATGTAAACAGCCACTGGGGTACAGGCACCGAAGACCAACATC CACAGGCTAACACCCCTTCAGTCCACACAAAGAAGCTTCATATTTTTTTATAAGCAT AGAAATAAAAACCAAGCCAATATTTGTGACTTTGCTCTGCTACCTGCTGTATTTATTA TATGAAGCATCTAAGTACTGTGAGGATGGGGTCTTCTCATTTGAGGCGTTAGGAT GTTGCTTTCTTTTCCATTAGTTAAACATTTTTTTCTCCTTTGGAGGAAGGGAATGAA ACATTTATGGCCTCAAGATACTATACATTTAAAGCACCCCAATGTCTCTCTTTTTTTT TTTTACTTCCCTTTCTTCTTCTTATATAACATGAAGAACATGTATTAATCTGATT TTTAAAGATCTTTTGTATGTTACGTGTTAAGGGCTGTTTGGTATCCCACTGAAATG TTCTGTGTTGCAGACCAGAGTCTGTTTATGTGAGGGGATGGGGCCATTGCATCCTTA GCCATTGTACAAAATATGTGGAGTAGTAACTTAATATGTAAAGTTGTAACATACATA CATTTAAATGGAATGCAGAAAGCTGTGAAATGTCTGTGTCTTATGTTCTCTGTAT TTATGCAGCTGATTTGTCTGTCTGTAAGTGTGGGTCCAAGGACTCCTAACTAC TTTGCATCTGTAATCCACAAAGATTCTGGGCAGCTGCCACCTCAGTCTCTTCTCTGTA TTATCATAGTCTGGTTTAAATAAACTATATAGTAAACAAAAAAGAAAAAAGAAAA AAA AAAAAAAAAAAAAAAAAAAAA		
	ORF Start: ATG at 141           ORF Stop: TAA at 3444		
	SEQ ID NO: 226	1101 aa	MW at 120838.0kD
NOV15a, CG142427-01 Protein Sequence	MSAKAISEQTGKELLYKFICTTSAIQNRKFYARVTPDWDWARLLQDHPWLLSQNLVVK PDQLIKRRGKGLVGVNLTLDGVKSWLKPRLGQEATVGKATGFLKNFLIEPFVPHSQA EEFYVCIYATREGDYVLFHHEGGVDVGDVDAKAQKLLVGVDEKLNPEDIKHLVLVHAP EDKKEILASFISGLNFYEDLYFTYLEINPLVVTKDGVYVLDLAAKV DATADYICKVK WGDIEFPFPFGREAYPEEAYIADLDAKSGASLKLTLNPKGRIWTMVAGGASVYYS TICDLGGVNELANYGEYS GAPSEQQTYDYAKTILSLMTRKHDPGKILIIIGGSIANFT NVAATFKGIVRAIRDYQGPLKEHEVTIFVRRGGPNYQGLRVMGEVKGTTGPIHVFG TETHMTAIVGMALGHRPIPNQPPTAAHTANFLLNASGSTSPAPSRSTASFESRADEV APAKKAKPAMPQDSVPSRSLQGKSTTLFSRHTKAIWGMQTRAVQGM LDFDYVCSR EPSVAAMVYPFTGDHKQKFWGHKEILIPVFKNMADAMRKHPVDVLINFA SLRSAYD STMETMNYAQIRTI AIIAEGIP EALTRKLIKADQKGVTTIIGPATVGGIKPGCFKIGN TGGMLDNILASKLYRPGSVAYVSRSGMSNELNNIISRTTDGVYEGVAIGGDRYPGST FMDHVLRYQDTPGVKMI VVLGEIIGTEYKICRGIKEGRLTPIVCWCIGTCATMFSS EVQFGHAGACANQASETAVAKNQALKEAGVFVPRSFDELGEI IQSVYEDLVANGVIVP AQEVPPPTVPM DYSWARELGLIRKPASFMTSICDERGQELIYAGMPITEVFKEEMGIG GVLGLLWFQKRLPKYSCQFIEMCLMVTADHGPAVSGAHNTIICARAGKDLVSSLTSG LTIGDRFGGALDAAAKMFSKAFDSGIIPMEFVNKMKKEGKLIMGIGHRVKSINNPDMR VQILKDYVRQHFPATPLLDYALEVEKITTSKKPNLILNVDGLIGVAFVDMLRNCGSFT REEADEYIDIGALNGIFVLGRSMGFIGHYLDQKRLKQGLYRHPWDDI SYVLP EHMMSM		
	SEQ ID NO: 227	4427 bp	
NOV15b,	GGCACGAGGCCGGGACAAAAGCCGGATCCCGGGAAGCTACCGGCTGCTGGGGTGCTCC		

CG142427-01  
DNA Sequence

GGATTTTGC GGGGTT CGTCGGGCTGTGGAAGAAGCGCCGCGCACGGACTTCGGCAGA  
GGTAGAGCAGGTCTCTCTGCAGCCATGTCCGCCAAGGCAATTTTCAGAGCAGACGGGCA  
AAGAACTCCTTTACAAGTTCATCTGTACCACCTCAGCCATCCAGAATCGGTTCAAGTA  
TGCTCGGGTCACTCCTGACACAGACTGGGCCCGCTTGCTGCAGGACCACCCCTGGCTG  
CTCAGCCAGAACTTGGTAGTCAAGCCAGACCAGCTGATCAAACGTCGTGGAAAACCTTG  
GTCTCGTTGGGGTCAACCTCACTCTGGATGGGGTCAAGTCTGGCTGAAGCCACGGCT  
GGGACAGGAAGCCACAGTTGGCAAGGCCACAGGCTTCCTCAAGAATTTCTGATCGAG  
CCCTTCGTCCCCCAGTCAAGCTGAGGAGTTCTATGTCTGCATCTATGCCACCCGAG  
AAGGGGACTACGTCCTGTTCACCACGAGGGGGGTGTGGACGTGGGTGATGTGGACGC  
CAAGGCCCAGAAGCTGCTTGTGGCGTGGATGAGAACTGAATCTGAGGACATCAAA  
AAACACCTGTTGGTCCACGCCCCCTGAAGACAAGAAAAGAAATTTCTGGCCAGTTTATCT  
CCGGCCTCTTCAATTTCTACGAGGACTTGTACTTCACCTACCTCGAGATCAATCCCCCT  
TGTAAGTACCAAAGATGGAGTCTATGTCTTGAATTTGGCGGCCAAGGTGGACGCCACT  
GCCGACTACATCTGCAAAGTGAAGTGGGGTGACATCGAGTTCCCTCCCCCTTCGGGC  
GGGAGGCATATCCAGAGGAAGCCTACATTGCAGACCTCGATGCCAAAAGTGGGGCAAG  
CCTGAAGCTGACCTTGCTGAACCCCAAAGGGAGGATCTGGACCATGGTGGCCGGGGT  
GGCGCCTCTGTCTGTACAGCGATACCATCTGTGATCTAGGGGGTGTCAACGAGCTGG  
CAAATATGGGGAGTACTCAGGCGCCCCAGCGAGCAGCAGACCTATGACTATGCCAA  
GACTATCCTCTCCCTCATGACCCGAGAGAAGCACCAGATGGCAAGATCCTCATCATT  
GGAGGCAGCATCGAAATTCACCAACGTGGCTGCCACGTTCAAGGGCATCGTGAGAG  
CAATTGAGATTACCAAGGGCCCCCTGAAGGAGCAGCAAGTCAAACTCTTGTCCGAAG  
AGGTGGCCCCAATATCAGGAGGGCTTACGGGTGATGGGAGAAGTCGGGAAGACCACT  
GGGATCCCCATCCATGTCTTGGCACAGAGACTCACATGACGGCCATTGTGGGCATGG  
CCCTGGGGCACCAGCCCCATCCCCAACCAGCCACCCACAGCGGCCCACTGCAAACTT  
CCTCCTCAACGCCAGCGGGAGCACATCGACGCCAGCCCCAGCAGGACAGCATCTTTT  
TCTGAGTCCAGGGCCGATGAGGTGGCGCTGCAAAGAAGGCCAAGCCTGCCATGCCAC  
AAGATTGAGTCCCAAGTCCAAGATCCCTGCAAGGAAGAGCACCACCTCTTCAGCCG  
CCACACCAAGGCCATTGTGTGGGGCATGCAGACCCGGGCCGTGCAAGGCATGCTGGAC  
TTTGACTATGTCTGCTCCCGAGACGAGCCCTCAGTGGCTGCCATGGTCTACCTTTCA  
CTGGGGACCAAGCAGAAGTTTACTGGGGGCACAAAGAGATCCTGATCCCTGTCTT  
CAAGAACATGGCTGATGCCATGAGGAAGCATCCGGAGGTAGATGTGCTCATCAACTTT  
GCCTCTCTCCGCTCTGCCTATGACAGCACCATGGAGACCATGAACTATGCCCAGATCC  
GGACCATCGCCATCATAGCTGAAGGCATCCCTGAGGCCCTCACGAGAAAGCTGATCAA  
GAAGGCGGACCAGAAGGGAGTGACCATCATCGGACCTGCCACTGTGAGGCATCAAG  
CCTGGGTGCTTTAAGATTGGCAACACAGGTGGGATGCTGGACAACATCCTGGCCTCCA  
AACTGTACCGCCAGGCAGCGTGGCCTATGTCTCAGTTCCGGAGGCATGTCCAACGA  
GCTCAACAATATCATCTCTCGGACCACGGATGGCGTCTATGAGGGCGTGGCCATTGGT  
GGGGACAGGTACCCGGGCTCCACATTCATGGATCATGTGTTACGCTATCAGGACACTC  
CAGGAGTCAAAATGATTGTGGTTCTTGGAGAGATTGGGGGCACTGAGGAATATAAGAT  
TTGCCGGGGCATCAAGGAGGGCCGCTCACTAAGCCATCGTCTGCTGGTGCATCGGG  
ACGTGTGCCACCATGTTCTCTCTGAGGTCCAGTTTGGCCATGCTGGAGCTTGTGCCA  
ACCAGGCTTCTGAACTGCAGTAGCCAAGAACCAGGCTTTGAAGGAAGCAGGAGTGT  
TGTGCCCCGGAGCTTTGATGAGCTTGGAGAGATCATCCAGTCTGTATACGAAGATCTC  
GTGGCCAATGGAGTCATTGTACCTGCCAGGAGGTGCCGCCCCCAACCGTGCCCATGG  
ACTACTCCTGGCCAGGAGCTTGGTTGATCCGCAACCTGCCCTCGTTTCATGACCAG  
CATCTGCGATGAGCGAGGACAGGAGCTCATCTACGCGGGCATGCCCATCACTGAGGTC  
TTCAAGGAAGAGATGGGCATTGGCGGGGCTCTCGGCCTCTCTGGTTCAGAAAAGGT  
TGCTTAAGTACTCTTGCCAGTTTCATTGAGATGTGTCTGATGTGACAGCTGATCAGG  
GCCAGCCGTCTCTGGAGCCACAAACACCATCATTTGTGCGCGAGCTGGGAAAGACCTG  
GTCTCCAGCTCACCTCGGGGCTGCTCACCATCGGGGATCGGTTTGGGGTGCCTTGG  
ATGCAGCAGCCAAGATGTTCAAGTAAAGCCTTTGACAGTGGCATTATCCCCATGGAGTT  
TGTGAACAAGATGAAGAAGGAAGGGAAGCTGATCATGGGCATTGGTCACCGAGTGAAG  
TCGATAAACAACCCAGACATGCGAGTGCAGATCCTCAAAGATTACGTACGGCAGCACT  
TCCCTGCCACTCCTCTGCTGATTATGCACTGGGAAGTAGAGAAGATTACCACTCGAA  
GAAGCCAAATCTTATCCTGAATGTAGATGGTCTCATCGGAGTCGATTTGTAGACATG  
CTTAGAACTGTGGGTCTTTACTCGGGAGGAAGCTGATGAATATATTGACATTGGAG  
CCCTCAATGGCATCTTTGTGCTGGGAAGGAGTATGGGGTTCATTGGACACTATCTTGA  
TCAGAAGAGGCTGAAGCAGGGGCTGTATCGTCATCCGTGGGATGATATTTTCATATGTT  
CTCCGGAAACACATGAGCATGTAACAGAGCCAGGAACCTACTGCAAGTAACTGAAGA  
CAAGATCTCTTCCCCAAGAAAAGTGTACAGACAGCTGGCAGTGGAGCCTGCTTTAT  
TTAGCAGGGGCTGGAATGTAACAGCCACTGGGGTACAGGCACCGAAGACCAACATC

	CACAGGCTAACACCCCTTCAGTCCACACAAAGAAGCTTCATATTTTTTTTATAAGCAT AGAAATAAAAACCAAGCCAATATTTGTGACTTTGCTCTGCTACCTGCTGATTTATTA TATGGAAGCATCTAAGTACTGTCAGGATGGGGTCTTCTCATTGTAGGGCGTTAGGAT GTTGCTTTCTTTTCCATTAGTTAAACATTTTTTCTCCTTTGGAGGAAGGGAATGAA ACATTTATGGCCTCAAGATACTATACATTTAAAGCACCCCAATGTCTCTCTTTTTTTT TTTTTACTTCCCTTCTTCTTCTTATATAACATGAAGAACATTGTATTAATCTGATT TTTAAAGATCTTTTGTATGTTACGTGTTAAGGGCTTGTTTGGTATCCCACTGAAATG TTCTGTGTTGCAGACCAGAGTCTGTTTATGTCAGGGGGATGGGGCCATTGCATCCTTA GCCATTGTCACAAAATATGTGGAGTAGTAACCTTAATATGTAAAGTTGTAACATACATA CATTTAAAATGGAAATGCAGAAAGCTGTGAAATGTCTTGTGTCTTATGTTCTCTGTAT TTATGCAGCTGATTGTCTGTCTGTAAGTGAAGTGTGGGTCCAAGGACTCCTAACTAC TTTGATCTGTAAATCCACAAAGATTCTGGGCAGCTGCCACCTCAGTCTCTTCTGTGA TTATCATAGTCTGGTTTAAATAAACTATATAGTAACAAAAAAAAAAAAAAAAAAAAA AA AAAAAAAAAAAAAAAAAAAA		
	ORF Start: ATG at 141		ORF Stop: TAA at 3444
	SEQ ID NO: 228	1101 aa	MW at 120838.0kD
NOV15b, CG142427-01 Protein Sequence	MSAKAISEQTGKELLYKFICTTSAIQNRFKYARVTPDWDARLLQDHPWLLSQNLVVK PDQLIKRRGKGLGVGNLTLDGVKSWLKPRLGQEAIVGKATGFLKNFLIEPFPVHSQA EEFYVCIYATREGDYVLFHHEGGVDVGDVDAKAQKLLVGVDEKLNPEDIKKHLLVHAP EDKKEILASFI SGLNFYEDLYFTYLEINPLVVTKDGVYVLDLAAKVDATADYICKVK WGDIEFPFPFGREAYPEEAYIADLDAKSGASLKLTLNPKGRIWTMVAGGASVYVSD TICDLGGVNELANYGEYS GAPSEQQTYDYAKTILSLMTRKHPDGKILIIIGGSIANFT NVAATFKGIVRAIRDYQGPLKEHEVTIFVRRGGPNYQEGLRVMGEVGGKTTGPIHVFV TETHMTAIVGMALGHRPINQPPPTAHTANFLLNASGSTSTPAPSRTASFSESRADEV APAKKAKPAMPQDSVPSRSLQKSTTLFSRHTKAIWGMQTRAVQGMDFDVFVCSR EPSVAAMVYPFTGDHKQKFYWGHEKILIPVFKNMADAMRKHPVDVLINFASLSAYD STMETMNYAQIRTAIAIEGIPALTRKLIKADQKGVTIIGPATVGGIKPGCFKIGN TGGMLDNILASKLYRPGSVAYVSRSGMSNELNNIISRTDGVYEGVAIGGDYRPGST FMDHVLRYQDTPGVKMIIVLGEIGGTEYKICRGIKEGRLTKPIVCWCIGTCATMFSS EVQFGHAGACANQASETAVAKNQALKEAGVFVPRSFDELGEIIQSVYEDLVANGVIVP AQEVPPPTVPM DYSWARELGLIRKPASFMTSICDERGQELIYAGMPITEVFKEEMGIG GVLGLLWFQKRLPKYSCQFIEMCLMVTADHGPVSGAHNTIICARAGKDLVSSLTSG LTI GDRFGGALDAAAKMFSKAFDSGII PMEFVNKMKEGKLMGIGHRVKSI NNPDMM VQILKDYVRQHF PATPLLDYALEVEKITTSKKPNLILNVDGLIGVAFVDMRLNCGSFT REEADEYIDIGALNGIFVLGRSMGFIGHYLDQKRLKQGLYRHPWDDISYVLPEHMSM		
	SEQ ID NO: 229	3238 bp	
NOV15c, CG142427-04 DNA Sequence	CCAGAATTCCACCATGTCCGGCCAAGGCAATTTTCAGAGCAGACGGGCAAGAAGTCTCTT TACAAGTTTCATCTGTACCACTCAGCCATCCAGAATCGGTTCAAGTATGCTCGGGTCA CTCCTGACACAGACTGGGCGCGCTTGCTGCAGGACCACCCCTGGTGTCTCAGCCAGAA CTTGGTAGTCAAGCCAGACCAGCTGATCAAACGTCGTGGAAAAGTCTGGTCTCGTTGGG GTCAACCTCACTCTGGATGGGGTCAAGTCTGGCTGAAGCCACGGCTGGGACAGGAAG CCACAGTTGGCAAGGCCACAGGCTTCTCAAGAACTTTCTGATCGAGCCCTTCGTCCC CCACAGTCAGGCTGAGGAGTTCTATGTCTGCATCTATGCCACCCGAGAAGGGGACTAC GTCCTGTTCCACCACGAGGGGGTGTGGACGTGGGTGATGTGGACGCCAAGGCCAGAA AGCTGCTTGTGGCGTGGATGAGAACTGAATCCTGAGGACATCAAAAAACCTGTT GGTCCACGCCCCTGAAGACAAGAAAGAAATTTCTGGCCAGTTTTATCTCCGGCCTCTTC AATTTCTACGAGGACTTGTACTTCACTACCTCGAGATCAATCCCCTTGTAAGTACCA AAGATGGAGTCTATGTCTTGACTTGGCGGCCAAGGTGGACGCCACTGCCGACTACAT CTGCAAAGTGAAGTGGGGTGACATCGAGTTCCCTCCCCCTTCGGGCGGGAGGCATAT CCAGAGGAAGCCTACATTGCAGGCCTCGATGCCAAAAGTGGGGCAAGCCTGAAGCTGA CCTTGCTGAACCCCAAAGGGAGGATCTGGACCATGGTGGCCGGGGTGGCCCTCTGT CGTGACAGCGATACCATCTGTGATCTAGGGGGTGTCAACGAGCTGGCAAACTATGGG GAGTACTCAGGCGCCCCAGCGAGCAGCAGACCTATGACTATGCAAGGACTATCTCTCT CCCTCATGACCCGAGAGAAGCACCCAGATGGCAAGATCCTCATCATTGGAGGCAGCAT CGAAACTTACCAACGTGGCTGCCACGTTCAAGGGCATCGTGAGAGCAATTCGAGAT TACCAGGGCCCCCTGAAGGAGCACGAAGTCACAATCTTTGTCCGAAGAGGTGGCCCCA ACTATCAGGAGGGCTTACGGGTGATGGGAGAAGTCGGGAAGACCCTGGGATCCCCAT		

	CCATGTCTTTGGCACAGAGACCCACACTGCAAACCTTCCTCCTCAACGCCAGCGGGAGC ACATCGACGCCAGCCCCAGCAGGACAGCATCTTTTCTGAGTCCAGGGCCGATGAGG TGGCGCCTGCAAAGAAGGCCAAGCCTGCCATGCCACAAGGAAAGAGCACCACCTCTT CAGCCGCCACACCAAGGCCATTGTGTGGGGCATGCAGACCCGGGCGGTGCAAGGCATG CTGGACTTTGACTATGTCTGCTCCCGAGACGAGCCCTCAGTGGCTGCCATGGTCTACC CTTTCACCTGGGGACCACAAGCAGAAGTTTACTGGGGGCACAAAGAGATCCTGATCCC TGTCTTCAAGAACATGGCTGATGCCATGAGGAAGCATCCGGAGGTAGATGTGCTCATC AACTTTGCCTCTCTCCGCTCTGCCTATGACAGCACCATGGAGACCACGAACTATGCC AGATCCGACCATCGCCATCATAGCTGAAGGCATTCCTGAGGCCCTCACGAGAAAGCT GATCAAGAAGGCGGACCAGAAGGGAGTGACCATCATCGGACCTGCCACTGTTGGAGGC ATCAAGCCTGGGTGCTTTAAGATTGGCAACACAGGTGGGATGCTGGACAACATCCTGG CCTCCAACTGTACCGCCAGGCAGCGTGGCCTATGCCTCACGTTCCGGAGGCATGTC CAACGAGCTCAACAATATCATCTCTCGGACCACGGATGGCGTCTATGAGGGCGTGGCC ATTGGTGGGGACAGGTACCCGGGCTCCACATTCATGGATCATGTGTTACGCTATCAGG ACACTCCAGGAGTCAAAATGATTGTGGTTCTTGGAGAGATTGGGGGCATGAGGAATA TAAGATTGCGGGGCATCAAGGAGGGCGCCTCACTAAGCCCATCGTCTGCTGGTGC ATCGGACGTGTGCCACCATGTTCTCCTCTGAGGTCCAGTTTGGCCATGCTGGAGCTT GTGCCAACCAGGCTTCTGAACTGCAGTAGCCAAGAACCAGGCTTTGAAGGAAGCAGG AGTGTGTTGTGCCCCGAGCTTTGATGAGCTTGGAGAGATCATCCAGTCTGTATACGAA GATCTCGTGGCCAAATGGAGTCATTGTACCTGCCAGGAGGTGCCGCCCCCAACCGTGC CCATGGACTACTCCTGGGCCAGGAGCTTGGTTTATCCGCAACCTGCCTCGTTCAT GACCAGCATCTGCGATGAGCGAGGACAGGAGCTCATCTACGGCGGCATGCCCATCACT GAGGTCTTCAAGGAAGAGATGGGCATTGGCGGGTCTCGCCCTCTCTGTTCCAGA AAAGGTTGCCTAAGTACTCTTGCCAGTTCATTGAGATGTGCTGATGGTGACAGCTGA TCACGGGCCAGCCGTCTCTGGAGCCCAACACCATCATTTGTGCGCGAGCTGGGAAA GACCTGGTCTCCAGCCTCACCTCGGGGCTGCTCACCATCGGGGATCGGTTTGGGGGTG CCTTGGATGCAGCAGCCAAGATGTTCACTAAAGCCTTTGACAGTGGCATTATCCCCAT GGAGTTTGTGAACAAGATGAAGAAGGAAGGAAGCTGATCATGGGCATTGGTCAACGA GTGAAGTCGATAAACAACCCAGACATGCGAGTGCAGATCCTCAAAGATTACGTACGGC AGCACTTCCCTGCCACTCCTCTGCTCGATTATGCACTGGAAGTAGAGAAGATTACCAC CTCGAAGAAGCCAATCTTATCCTGAATGTAGATGGTCTCATCGGAGTCGCATTTGTA GACATGCTTAGAACTGTGGTCTCTTACTCGGGAGGAAGCTGATGAATATATTGACA TTGGAGCCCTCAATGGCATCTTTGTGCTGGGAAGGAGTATGGGGTTCATTGGACACTA TCTTGATCAGAAGAGGCTGAAGCAGGGGCTGTATCGTCATCCGTGGGATGATATTTCA TATGTTCTTCCGGAACACATGAGCATGTAAGCGGCCGCTTTTTCCTT		
	ORF Start: at 2		ORF Stop: TAA at 3218
	SEQ ID NO: 230	1072 aa	MW at 117722.3kD
NOV15c, CG142427-04 Protein Sequence	QNSTMSAKAISEQTGKELLYKFICTTSAIQNRFKYARVTPDWDWARLLQDHPWLLSQN LVVKPDQLIKRRGKLGVLGVNLTLDGVKSWLKPRLGQEATVGKATGFLKNFLIEPFPV HSQAEEFYVCIYATREGDYVLFHHEGGVDVGDVDAKAQKLLVGVDEKLNPEDIKKHLL VHAPEDKKEILASFISGLNFYEDLYFTYLEINPLVVTKDGVYVLDLAAKV DATADYI CKVKWGDIEFPFPFGREAYPEEAYIAGLDAKSGASLKLTLNPKGRIWTMVAGGGASV VYSDTICDLGGVNELANYGEYS GAPSEQQTYDYAKTILSLMTRKHDPGKILIIIGGSI ANFTNVAATFKGIVRAIRDYQGPKHEVTFVRRGGPNYQEGLRVMGEVKGTTGIPI HVFGTETHANFLLNASGSTSTPAPSRTASFSESRADEVAPAKKAKPAMPQGKSTTLF SRHTKAI VWGMQTRAVQGMDFDYVCSRDEPSVAAMVYPFTGDHKQKFYWGHEKILIP VFKNMADAMRKHPEVDVLINFASLR SAYDSTMETTNYAQIRTI AIIAEGIPEALTRKL IKKADQKGVTTIIGPATVGGIKPGCFKIGNTGMLDNILASKLYRPGSVAYASRSGGMS NELNNIISRTTDGVYEGVAIGGDRYPGSTFMDHVLRYQDTPGVKMI VVLGEIGGTBEY KICRGIKEGRLTKPIVCWICIGTCATMFSSEVQFHAGACANQASETAVAKNQALKEAG VFVPRSFDELGEIIQSVYEDLVANGVIVPAQEVPPPTVPMDYSWARELGLIRKPASFM TSICDERGQELIYAGMPITEVFKEEMGIGVLGLLWFQKRLPKYSCQFIEMCLMVTAD HGPAVSGAHNTIICARAGKDLVSSLTSGLLTIGDRFGGALDAAAKMFSKAFDSGIIPM EFVNKMKKEGKLIMGIGHRVKSINNPDMRVQILKDYVRQHF PATPLLDYALEVEKITTT SKKPNLILNVDGLIGVAFVDMLRNCGSF TREADEYIDIGALNGIFVLGRSMGFIGHY LDQKRLKQGLYRHPWDDISYVLP EHMMSM		
	SEQ ID NO: 231	3307 bp	
NOV15d,	CCAGAATTCACCATGTCCGCCAAGGCAATTCAGAGCAGACGGGCAAAGAACTCCTT		

CG142427-02 DNA Sequence	<p>TACAAGTTCATCTGTACCACCTCAGCCATCCAGAATCGGTTCAAGTATGCTCGGGTCA  CTCCTGACACAGACTGGGCCCCGCTTGCTGCAGGACCAACCCCTGGCTGCTCAGCCAGAA  CTTGGTAGTCAAGCCAGACCAGCTGATCAAACGTCGTGGAAAACCTTGGTCTCGTTGGG  GTCAACCTCACTCTGGATGGGGTCAAGTCTGGCTGAAGCCACGGCTGGGACAGGAAG  CCACAGTTGGCAAGGCCACAGGCTTCCTCAAGAACTTTCTGATCGAGCCCTTCGTCCC  CCACAGTCAGGCTGAGGAGTTCTATGTCTGCATCTATGCCACCCGAGAAGGGGACTAC  GTCCTGTTCCACCACGAGGGGGGTGTGGACGTGGGTGATGTGGACGCCAAGGCCCAGA  AGCTGCTTGTGGCGTGGATGAGAACTGAATCCTGAGGACATCAAAAAACACCTGTT  GGTCCACGCCCTGAAGACAAGAAAGAAATTCTGGCCAGTTTATCTCCGCCCTCTTC  AATTTCTACGAGGACTTGTACTTCACCTACCTCGAGATCAATCCCCCTTGTAGTGACCA  AAGATGGAGTCTATGTCTTGACTTGGCGGCCAAGGTGGACGCCACTGCCGACTACAT  CTGCAAAGTGAAGTGGGGTGACATCGAGTTCCCTCCCCCTTCGGGCGGGAGGCATAT  CCAGAGGAAGCCTACATTGCAGACCTCGATGCCAAAAGTGGGGCAAGCCTGAAGCTGA  CCTTGCTGAACCCCAAAGGGAGGATCTGGACCATGGTGGCCGGGGTGGCGCCTCTGT  CGTGACAGCGATACCATCTGTGATCTAGGGGGTGTCAACGAGCTGGCAAACATATGGG  GAGTACTCAGGCGCCCCCAGCGAGCAGACCTATGACTATGCCAAGACTATCCTCT  CCCTCATGACCCGAGAGAAGCACCCAGATGGCAAGATCCTCATATTGGAGGCAGCAT  CGCAAACCTCACCAACGTGGCTGCCACGTTCAAGGGCATCGTGAGAGCAATTTCGAGAT  TACCAGGGCCCCCTGAAGGAGCACGAAGTCACAATCTTTGTCCGAAGAGGTGGCCCCA  ACTATCAGGAGGGCTTACGGGTGATGGGAGAAGTCGGGAAGACCACTGGGATCCCCAT  CCATGTCTTTGGCACAGAGACTCACATGACGGCCATTGTGGGCATGGCCCTGGGCCAC  CGCCCCATCCCCAACAGCCACCCACAGCGGCCACACTGCAAACTTCCTCCTCAACG  CCAGCGGGAGCACATCGACGCCAGCCCCAGCAGGACAGCATCTTTTCTGAGTCCAG  GGCCGATGAGGTGGCGCTGCAAAGAAGGCCAAGCCTGCCATGCCACAAGGAAAGAGC  ACCACCCCTCTCAGCCGCCACACCAAGGCCATTGTGTGGGGCATGCAGACCCGGGCCG  TGCAAGGCATGCTGGACTTTGACTATGTCTGCTCCCGAGACGAGCCCTCAGTGGCTGC  CATGGTCTACCCCTTCACTGGGGACCACAAGCAGAAGTTTACTGGGGGCACAAAGAG  ATCCTGATCCCTGTCTTCAAGAACATGGCTGATGCCATGAGGAAGCACCCGAGGTAG  ATGTGCTCATCAACTTTGCCTCTCTCCGCTCTGCCTATGACAGCACCATGGAGACCAT  GAACTATGCCCAGATCCGGACCATCGCCATCATAGCTGAAGGCATCCCTGAGGCCCTC  ACGAGAAAGCTGATCAAGAAGGCGGACCAGAAGGGAGTGACCATCATCGGACCTGCCA  CTGTTGGAGGCATCAAGCCTGGGTGCTTTAAGATTGGCAACACAGGTGGGATGCTGGA  CAACATCTGGCCTCCAAACTGTACCGCCAGGCAGCGTGGCCTATGTCTCACGTTCC  GGAGGCATGTCCAACGAGCTCAACAATATCATCTCTCGGACCACGGATGGCGTCTATG  AGGGCGTGGCCATTGTTGGGGACAGGTACCCGGCTCCACATTCATGGATCATGTGTT  ACGCTATCAGGACACTCCAGGAGTCAAAATGATTTGTGGTTCTTGGAGAGATTGGGGGC  ACTGAGGAATATAAGATTGCGGGGGCATCAAGGAGGGCCGCTCCTAAGCCCCATCG  TCTGCTGGTGCATCGGGACGTGTGCCACCATGTTCTCTCTGAGGTCCAGTTTGGCCA  TGCTGGAGCTTGTGCCAACAGGCTTCTGAACTGCAGTAGCCAAGAACCAGGCTTTG  AAGGAAGCAGGAGTGTGTTGTCGCCGGAGCTTTGATGAGCTTGGAGAGATCATCCAGT  CTGTATACGAAGATCTCGTGGCCAATGGAGTCATTGTACCTGCCAGGAGGTGCCGCC  CCCAACCGTGCCCATGGACTACTCCTGGGCCAGGGAGCTTGGTTGATCCGCAAACCT  GCCTCGTTCATGACCAGCATCTGCGATGAGCGAGGACAGGAGCTCATCTACGCGGGCA  TGCCCATCACTGAGGTCTTCAAGGAAGAGATGGGCATTGGCGGGGTCTCGGCCTCCT  CTGGTTCCAGAAAAGGTTGCCTAAGTACTCTTGCCAGTTCATTGAGATGTGTCTGATG  GTGACAGCTGATCACGGGCCAGCCGTCTCTGGAGCCACAACACCATCATTTGTGCCG  GAGCTGGGAAAGACCTGGTCTCCAGCCTCACCTCGGGGTGCTCACCATCGGGGATCG  GTTTGGGGGTGCTTGGATGCAGCAGCCAAGATGTTCAAGTAAAGCCTTTGACAGTGGC  ATTATCCCATGGAGTTTGTGAACAAGATGAAGAAGGAAGGAAGCTGATCATGGGCA  TTGGTCACCGAGTGAAGTCGATAAACAACCCAGACATGCGAGTGCAGATCCTCAAAGA  TTACGTCAGGCAGCACTTCCCTGCCACTCCTCTGCTCGATTATGACTGGAAGTAGAG  AAGATTACCACCTCGAAGAAGCCAAATCTTATCCTGAATGTAGATGTTCTCATCGGAG  TCGCATTGTAGACATGCTTAGAACTGTGGGTCTTTACTCGGGAGGAAGCTGATGA  ATATATTGACATTGGAGCCCTCAATGGCATCTTTGTGCTGGGAAGGAGTATGGGGTTC  ATTGGACACTATCTTGATCAGAAGAGGCTGAAGCAGGGGCTGTATCGTCATCCGTGG  ATGATATTCATATGTTCTTCCGGAACACATGAGCATGTAAGCGGCCCTTTTTCTCT  T</p>
	<div>ORF Start: at 2</div> <div>ORF Stop: TAA at 3287</div>
	<div>SEQ ID NO: 232</div> <div>1095 aa</div> <div>MW at 120201.2kD</div>

NOV15d, CG142427-02 Protein Sequence	<p>QNSTMSAKAISEQTGKELLYKFICTTSAIQNRFKYARVTPDTDWARLLQDHPWLLSQN  LVVKPDQLIKRRGKLGVLGVNLTLDGVKSWLKPRLGQEQATVGKATGFLKNFLIEPFVP  HSQAEEFYVCIYATREGDYVLFHHEGGVDVGDVDAKAQKLLVGVDEKLNPEDIKKHLL  VHAPEDKKEILASFISGLNFYEDLYFTYLEINPLVVTKDGVVYLDLAAKVDATADYI  CKVKWGDIEFPFPFGREAYPEEAYIADLDAKSGASLKLTLNPKGRIWTMVAGGASV  VYSDTICDLGGVNELANYGEYS GAPSEQTYDYAKTILSLMTREKHPDGKILIIIGSSI  ANFTNVAATFKGIVRAIRDYQGPLKEHEVTIFVRRGGPNYQEGLRVMGEVGTKTIGPI  HVFGETHMTAIVGMALGHRPIPNQPPTAAHTANFLNAGSGTSTPAPSRTASFSES  ADEVAPAKKAKPAMPQGKSTLFSRHTKAIWGMQTRAVQGMDFDYVCSRDEPSVAA  MVYPFTGDHKQKFYWGHEKILIPVFKNMADAMRKHPEDVVLINFASLRSAYDSTMETM  NYAQIRTI AIAEBIPEALTRKLIKADQKGVTTIIGPATVGGIKPGCFKIGNTGMLD  NILASKLYRPGSVAYVSRSGMSNELNNIISRTTDGVYEGVAIGGDYRPGSTFMDHVL  RYQDTPGVKMIIVLGEIGGT EYKICRGIKEGRLTKPIVCWCIGTCATMFSSEVQFGH  AGACANQASETAVAKNQALKEAGVFVPRSFDELGEIIQSVYEDLVANGVIVPAQEVPP  PTVPMDYWARELGLIRKPASFMTSICDERGQELIYAGMPITEVFKBEMGIGGVGLLL  WFQKRLPKYSCQFIEMCLMVTADHGPAVSGAHNTIICARAGKDLVSSLTSGLLTIGDR  FGGALDAAAKMFSKAFDSGLIPMEFVNKMKKEGKLIMGHVRVKSINPNMDRVQILKD  YVRQHFPATPLLDYALEVEKITTSKKPNLILNVDGLIGVAFVDMLRNCGSFTR EEADE  YIDIGALNGIFVLGRSMGFIGHYLDQKRLKQGLYRHPWDDISYVLP EHMMS</p>
	<div>SEQ ID NO: 233</div> <div>2290 bp</div>
NOV15e, CG142427-03 DNA Sequence	<p>CCAGAATTCACCATGTGCGCCAAGGCCAATTCAGAGCAGACGGGCAAAGAATCCTT  TACAAGTTCATCTGTACCACCTCAGCCATCCAGAATCGGTTCAAGTATGCTCGGGTCA  CTCCTGACACAGACTGGGCCCGCTTGCTGCAGGACCACCCCTGGCTGCTCAGCCAGAA  CTTGCTAGTCAAGCCAGACCAGCTGATCAAACGTCGTGGAAAACCTTGGTCTCGTTGGG  GTCAACCTCACTCTGGATGGGGTCAAGTCTGCTGGAAGCCACGGCTGGGACAGGAAG  CCACAGTGAGTGGGCATGGGGTCAAGATGAACGTGTGTGTAACAGAAGCAAAATATGG  TCACCTTCAGGTTGGCAAGGCCACAGGCTTCCTCAAGAACTTTCTGATCGAGCCCTTC  GTCCCCACAGTCAGGCTGAGGAGTTCTATGTCTGCATCTATGCCACCCGAGAAGGGG  ACTACGTCCTGTTCACACAGAGGGGGGTGTGGACGTGGGTGATGTGGACGCCAAGGC  CCAGAAGCTGCTTGTGGCGTGGATGAGAAAAGTGAATCCTGAGGACATCAAAAAACAC  CTGTTGGTCCACGCCCTGAAGACAAGAAAAGAAATTCGTGCCAGTTTATCTCCGGCC  TCTTCAATTTCTACGAGGACTTGTACTTACCTACCTCGAGATCAATCCCTTGTAGT  GACCAAAGATGGAGTCTATGTCTTGTACTTGGCGGCCAAGGTGGACGCCACTGCCGAC  TACATCTGCAAAGTGAAGTGGGGTGACATCGAGTTCCTCCCTCCCTTCGGGCGGGAGG  CATATCCAGAGGAAGCCTACATTGCAGACCTCGACGCCAAAAGTGGGGCAAGCCTGAA  GCTGACCTTGCTGAACCCCAAAGGGAGGATCTGGACCATGTTGGCCGGGGGTGGCGCC  TCTGTCTGTACAGCGATACCATCTGTGATCTAGGGGGTGTCAACCGAGCTGGCAAACT  ATGGGGAGTACTCAGGCGCCCCCAGCGAGCAGCAGACCTATGACTATGCCAAGACTAT  CCTCTCCCTCATGACCCGAGAGAAGCAGCCAGATGGCAAGATCCTCATCATTGAGGC  AGCATCGCAAACCTTACCAACGTGGCTGCCACGTTCAAGGGCATCGTGAGAGCAATTC  GAGATTACCAGGGCCCCCTGAAGGAGCAGCAAGTCACAATCTTTGTCCGAAGAGGTGG  CCCCAACTATCAGGAGGGCTTACGGGTGATGGGAGAAAGTCGGGAAGCACTGGGATG  CCCATCCATGTCTTTGGCACAGAGACTCACATGACGGCCATTGTGGGCATGGCCCTGG  GCCACCGGCCCATCCCCAACAGCCACCCACAGCGGCCACACTGCAAACCTCTCTCT  CAACGCCAGCGGGAGCACATCGACGCCAGCCCCCAGCAGGACAGCATCTTTTCTGAG  TCCAGGGCCGATGAGGTGGCGCCTGCAAGAAGGCCAAGCCTGCCATGCCACAAGGAA  AGAGCACCACCTCTTACGCCGCCACCAAGGCCATTGTGTGGGGCATGCAGACCCG  GGCGGTGCAAGGCATGTGGACTTTGACTATGTCTGCTCCCGAGAGCCCTCAAGTG  GCTGCCATGGTCTACCTTTCACTGGGGACCACAAGCAGAAGTTTACTGGGGGCACA  AAGAGATCCTGATCCCTGTCTTCAAGAACATGGCTGATGCCATGAGGAAGCACC CGGA  GGTAGATGTCTCATCACTTTGTCTCTCCGCTCTGCCTTGGATGCAGCAGCCAAG  ATGTTCAAGTAAAGCCTTTGACAGTGGCATTATCCCCATGGAGTTTGTGAACAAGATGA  AGAAGGAAGGAAGCTGATCATGGGCATTGGTCACCGAGTGAAGTGTGATAAACAACCC  AGACATGCGAGTGCGGATCTCAAAGATTACGTCAGGCAGCACTTCCCTGCCACTCCT  CTGCTCGATTATGCACTGGAAGTAGAGAAGATTACCACTCGAAGAAGCCAAATCTTA  TCCTGAATGTAGATGGTCTCATCGGAGTCGCATTGTAGACATGCTAGAACTGTGG  GTCCTTTACTCGGGAGGAAGCTGATGAATATATTGACATTGGAGCCCTCAATGGCATC  TTTGTGCTGGGAAGGAGTATGGGGTTCATTGGACACTATCTTGATCAGAAGAGGCTGA  AGCAGGGGCTGTATCGTATCCGTTGGATGATATTTCATATGTTCTTCCGGAACACAT  GAGCATGTAAGCGGCCGCTTTTTTCTCTT</p>

	ORF Start: at 2		ORF Stop: TAA at 2270
	SEQ ID NO: 234	756 aa	MW at 83890.7kD
NOV15e, CG142427-03 Protein Sequence	QNSTMSAKAISEQTGKELLYKFICTTSAIQNRFKYARVTPDPTDWARLLQDHPWLLSQN LUVKPDQLIKRRGKLGVLGNLTLDGVKSWLKPRLGQEAATVSGHVKMNVCGNRSKYG HLQVGKATGFLKNFLIEPFPVPHSQAEFFVYCIYATREGDYVLFHHEGGVDVGDVDAKA QKLLVGVDEKLNPEDIKKHLLVHAPEDKKEILASFI SGLFNFYEDLYFTYLEINPLVV TKDGVYVLDLAAKV DATADYICKVKWGDIEFPFPFGREAYPEEAYIADLDKSGASLK LTLNPNKGRIWTMVAGGASVVYSDTICDLGGVNELANYGEYS GAPSEQQTYDYAKTI LSLMTREKHPDGKILIIIGGSIANFTNVAATFKGI VRAIRDYQGPLEHEVTIFVRRGG PNYQEGLRVMGEVGKTTGPIHVFGEETHMTAIVGMALGHRPIPNQPPTAAHTANFLL NASGSTSTPAPSRTASFSESRADDEVAPAKKAKPAMPQGGKSTTLFSRHTKAI VWGMQTR AVQGMLDFDYVCSRDEPSVAAMVYPFTGDHKQKIFYWGHKEILLIPVFKNMADAMRKHPE VDVLINFASLR SALDAAAKMFSAFDSGII PMEFVNKMKKEGKLIMGIGHRVKSINNP DMRVRI LKDYVRQHFPATPLLDYALEVEKITTSKKNLILNVDGLIGVAFVDMLRNCG SFTREEADEYIDIGALNGIFVLGRSMGFIGHYLDQKRLKQGLYRHPWDDISYVLPPEHMSM		
	SEQ ID NO: 235	3317 bp	
NOV15f, 256388552 DNA Sequence	CCAGAATTCACCATGTGCGGCCAAGGCAATTCAGAGCAGACGGGCAAAGA ACTCCTT TACAAGTTCATCTGTACCACCTCAGCCATCCAGAATCGGTTCAAGTATGCTCGGGTCA CTCCTGACACAGACTGGGCCCGCTTGCTGCAGGACCACCCCTGGCTGCTCAGCCAGAA CTTGGTAGTCAAGCCAGACCAGCTGATCAAACGTCGTGGAAAACTTGGTCTCGTTGGG GTCAACCTCACTCTGGATGGGGTCAAGTCTTGGCTGAAGCCACGGCTGGGACAGGAAG CCACAGTTGGCAAGGCCACAGGCTTCCTCAAGAACTTTCTGATCGAGCCCTTCGTCCC CCACAGTCAGGCTGAGGAGTTCATATGCTGTCATCTATGCCACCCGAGAAGGGGACTAC GTCCTGTTCACCACGAGGGGGGTGTGGACGTGGGTGATGTGGACGCCAAGGCCAGAA AGCTGCTTGTGGCGTGGATGAGAACTGAATCCTGAGGACATCAAAAAACACCTGTT GGTCCACGCCCTGAAGACAAGAAAGAAATCTGGCCAGTTTTATCTCCGGCTCTTTC AATTTCTACGAGGACTTGTACTTCACCTACCTCGAGATCAATCCCCTTGTAGTGACCA AAGATGGAGTCTATGTCTTGACTTGGCGGCCAAGGTGGACGCCACTGCCGACTACAT CTGCAAAGTGAAGTGGGGTGACATCGAGTTCCTCCCCCTTCGGGCGGGAGGCATAT CCAGAGGAAGCCTACATTGCGAGCCTCGATGCCAAAAGTGGGGCAAGCTGAAGCTGA CCTTGCTGAACCCCAAAGGGAGGATCTGGACCATGGTGGCGGGGGTGGCGCCTCTGT CGTGTACAGCGATACCATCTGTGATCTAGGGGGTGTCAACGAGCTGGCAAACATATGGG GAGTACTCAGGCGCCCCCAGCGAGCAGCAGACCTATGATTATGCCAAGACTATCCTCT CCCTCATGACCCGAGAGAAGCACCAGATGGCAAGATCCTCATATTGGAGGCAGCAT CGCAAACCTTACCAACGTGGCTGCCACGTTCAAGGGCATCGTGAGAGCAATTCGAGAT TACCAGGGCCCCCTGAAGGAGCAGCAAGTCACAATCTTTGTCCGAAGAGGTGGCCCCA ACTATCAGGAGGGCTTACGGGTGATGGGAGAAGTCGGGAAGACCACTGGGATCCCCAT CCATGTCTTTGGCACAGAGACTCACATGACGGCCATTGTGGGCATGGCCCTGGGGCCAC CGGCCCATCCCCAACCAGCCACCCACAGCGGCCACACTGCAAACCTCCTCCTCAACG CCAGCGGGAGCACATCGACGCCAGCCCCAGCAGGACAGCATCTTTTTCTGAGTCCAG GGCCGATGAGGTGGCGCTGCAAAGAAGGCCAAGCCTGCCATGCCACAAGATTCAATC CCAAGTCCAAGATCCCTGCAAGGAAAGAGCACCACCCTCTTCAGCCGCCACACCAAGG CCATTGTGTGGGGCATGCAGACCCGGGCGGTGCAAGGCATGCTGGACTTTGACTATGT CTGCTCCCGAGACGAGCCCTCAGTGGCTGCCATGGTCTACCCCTTCTACTGGGGACCAC AAGCAGAAGTTTACTGGGGGCACAAAGAGATCCTGATCCCTGTCTTCAAGAACATGG CTGATGCCATGAGGAAGCACCAGGAGGTAGATGTGCTCATCAACTTGCCTCTCTCCG CTCTGCCTATGACAGCACCATGGAGACCATGAACATATGCCAGATCCGAGACCATCGCC ATCATAGCTGAAGGCATCCCTGAGGCCCTCACGAGAAAGCTGATCAAGAAGGCGGACC AGAAGGGAGTGACCATCATCGGACCTGCCACTGTTGGAGGCATCAAGCCTGGGTGCTT TAAGATTGGCAACACAGGTGGGATGCTGGACAACATCCTGGCCTCCAACTGTACCGC CCAGGCAGCGTGGCCTATGTCTCACGTTCCGGAGGCATGTCCAACGAGCTCAACAATA TCATCTCTCGGACCAAGGATGGCGTCTATGAGGGCGTGGCCATTGGTGGGACAGGTA CCCGGGCTCCACATTGATGATCATGTGTTACGCTATCAGGACACTCCAGGAGTCAAAA ATGATTGTGGTTCTTGGAGAGATTGGGGGCACTGAGGAATATAAGATTTCGGGGGCA TCAAGGAGGGGCGCCTCACTAAGCCCACGCTCTGCTGCTGTCATCGGGACGTGTGCCAC CATGTTCTCCTCTGAGGTCCAGTTTGGCCATGCTGGAGCTTGTGCCAACAGGCTTCT GAAACTGCAGTAGCCAAGAACCAGGCTTGAAGGAAGCAGGAGTGTGTTGTGCCCGGA GCTTTGATGAGCTTGGAGAGATCATCCAGTCTGTATACGAAGATCTCGTGGCCAATGG		

	AGTCATTGTACCTGCCCAGGAGGTGCGCCCCCAACCGTGCCCATGGACTACTCCTGG GCCAGGGAGCTTGGTTTGTATCCGCAAACCTGCCTCGTTTCATGACCAGCATCTGCGATG AGCGAGGACAGGAGCTCATCTACGCGGGCATGCCATCACTGAGGTCTTCAAGGAAGA GATGGGCATTTGGCGGGTCTCTCGGCTCCTCTGGTTCAGAAAAGGTTGCCTAAGTAC TCTTGCCAGTTTCATTGAGATGTGTCTGATGGTGACAGCTGATCACGGGCCAGCCGTCT CTGGAGCCACAAACCATCATTTGTGCGCGAGCTGGGAAAGACCTGGTCTCCAGCCT CACCTCGGGGCTGCTCACCATCGGGGATCGGTTTGGGGGTGCCTTGGATGCAGCAGCC AAGATGTTCACTAAAGCCTTTGACAGTGGCATTATCCCCATGGAGTTTGTGAACAAGA TGAAGAAGGAAGGAAGCTGATCATGGGCATTGGTCACCGAGTGAAGTCGATAAACA CCCAGACATGCGAGTGCAGATCCTCAAAGATTACGTGAGGCAGCACTTCCCTGCCACT CCTCTGCTCGATTATGCACTGGAAGTAGAGAAGATTACCACCTCGAAGAAGCCAAATC TTATCCTGAATGTAGATGGTCTCATCGGAGTCGCATTGTAGACATGCTTAGAACTG TGGGTCTTTACTCGGGAGGAAGCTGATGAATATATTGACATTGGAGCCCTCAATGGC ATCTTTGTGCTGGGAAGGAGTATGGGGTTCATTGGACACTATCTTGATCAGAAGAGGC TGAAGCAGGGGCTGTATCGTCATCCGTGGGATGATATTTTATATGTTCTCCGGAACA CATGAGCATGT		
	ORF Start: at 2		ORF Stop: end of sequence
	SEQ ID NO: 236	1106 aa	MW at 121268.4kD
NOV15f, 256388552 Protein Sequence	QNSTMSAKAISEQTGKELLYKFICTTSAIQNRFKYARVTPDTDWARLLQDHPWLLSQN LVVKPDQLIKRRGKLGVLGNLTLDGVKSWLKPRLGQEATVGKATGFLKNFLIEFPFV HSQAEEFYVCIYATREGDYVLFHHEGGVDVGDVDAKAQKLLVGVDEKLNPEDIKKHLL VHAPEDKKEILASFISGLNFYEDLYFTYLEINPLVVTKDG VYVLDLAAKV DATADYI CKVKWGDIEFPFPFGREAYPEEAYIADLDAKSGASLKLTLNPKGRIMTMVAGGASV VYSDTICDLGGVNELANYGEYS GAPSEQQTYDYAKTILSLMTRKHPDGKLIIGGSI ANFTNVAATFKGIVRAIRDYQGPLKEHEVTIFVRRGGPNYQEGLRVMGEVGKTTGIPI HVFGTETHMTAIVGMALGHRPIPNQPPTAHTANFLLNASGSTSTPAPSR TASFSER ADEVAPAKKAKPAMPQDSVPSRSLQKSTTLFSRHTKAI VWGMQTRAVQGM LDFDYV CSRDEPSVAAMVYPFTGDHKQFYWGHEILIPVFKNMADAMRKHPEDV LINFASLR SAYDSTMETMNYAQIRITAI IAEGIP EALTRKLIKADQKGVTIIGPATVGGIKPGCF KIGNTGGMLDNILASKLYRPGSVAYVSRSGGMSNELNNIISR TTDGVYEGVAIGGDRY PGSTFMDHVLRYQDTPGVKMIVVLGEIGTEEYKICRGIKEGRLTKPIVCWCIGTCAT MFSSEVQFGHAGACANQASE TAVAKNQALKEAGVFVPRS FDELGEIIQS VYEDLVANG VIVPAQEVPPPTVPM DYSWARELGLIRKPASFMTSICDERGQELIYAGMPI TEVFKEE MGIGGVLGLLWFQKRLPKYSCQFIEMCLMVTADHGPAVSGAHNTIICARAGKDLVSSL TSGLLTIGDRFGGALDAAAKMFSKAFDSGII PMEFVNKMKKEGKLIMGIGHRVKSINN PDMRVQLKDYVRQHF PATPLLDYALEVEKITTSKKPNLILNVDGLIGVAFVDM L RNC GSFTREEADEYIDIGALNGIFVLGRSMGFIGHYLDQKRLKQGLYRHPWDDISYVLPEH MSMX		
	SEQ ID NO: 237	3307 bp	
NOV15g, 256420210 DNA Sequence	CCAGAATTCACCATGTCGGCCAGGCAATTCAGAGCAGACGGGCAAGAACTCCTT TACAAGTTCATCTGTACCACCTCAGCCATCCAGAATCGGTTCAAGTATGCTCGGGTCA CTCCTGACACAGACTGGGCCCGCTTGCTGCAGGACCACCCCTGGCTGCTCAGCCAGAA CTTGGTAGTCAAGCCAGACCAGCTGATCAAACGTCGTGGAAGAACTTGGTCTCGTTGGG GTCAACCTCACTCTGGATGGGGTCAAGTCTCGGCTGAAGCCACGGCTGGGACAGGAAG CCACAGTTGGCAAGGCCACAGGCTTCTCAAGAACTTTCTGATCGAGCCCTTCGTCCC CCACAGTCAGGCTGAGGAGTTCTATGTCTGCATCTATGCCACCCGAGAAGGGGACTAC GTCCTGTTCCACCACGAGGGGGGTGTGGACGTGGGTGATGTGGACGCCAAGGCCCAGA AGCTGCTTGTGGCGTGGATGAGAACTGAATCCTGAGGACATCAAAAAACCTGTT GGTCCACGCCCCGTAAGACAAGAAAGAAATTCTGGCCAGTTTATCTCCGGCTCTTC AATTTCTACGAGGACTTGTACTTACCTACCTCGAGATCAATCCCCTTGTAGTGACCA AAGATGGAGTCTATGTCCTTGACTTGGCGGCCAAGGTGGACGCCACTGCCGACTACAT CTGCAAAGTGAAGTGGGGTGACATCGAGTTCCCTCCCCCTTCGGGCGGGAGGCATAT CCAGAGGAAGCCTACATTGCAGACCTCGATGCCAAAAGTGGGGCAAGCCTGAAGCTGA CCTTGCTGAACCCCAAAGGGAGGATCTGGACCATGGTGGCCGGGGGCTGGCCCTGTGT CGTGTAACAGGATACCATCTGTGATCTAGGGGGTGTCAACGAGCTGGCAAACTATGGG GAGTACTCAGGCGCCCCCAGCGAGCAGCAGACCTATGACTATGCCAAGACTATCCTCT CCCTCATGACCCGAGAGAAGCACCAGATGGCAAGATCCTCATCATTGGAGGCAGCAT		

	<p>CGCAAAC TTCACCAACGTGGCTGCCACGTTCAAGGGCATCGTGAGAGCAATTCGAGAT  TACCAGGGCCCCCTGAAGGAGCACGAAGTCACAATCTTTGTCCGAAGAGGTGGCCCCA  ACTATCAGGAGGGCTTACGGGTGATGGGAGAAGTCGGGAAGACCACTGGGATCCCCAT  CCATGTCTTTGGCACAGAGACTCACATGACGGCCATTGTGGGCATGGCCCTGGGGCCAC  CGGCCCATCCCCAACCCAGCCACCCACAGCGGCCACACTGCAAACTTCTCTCTCAACG  CCAGCGGGAGCACATCGACGCCAGCCCCAGCAGGACAGCATCTTTTCTGAGTCCAG  GGCCGATGAGGTGGCGCTGCAAAGAAGGCCAAGCCTGCCATGCCACAAGGAAAGAGC  ACCACCCTCTTCAGCCGCCACACCAAGGCCATTGTGTGGGGCATGCAGACCCGGGGCCG  TGCAAGGCATGCTGGACTTTGACTATGTCTGCTCCCGAGACGAGCCCTCAGTGGCTGC  CATGGTCTACCCTTTCACTGGGGACCACAAGCAGAAGTTTACTGGGGGCACAAAGAG  ATCCTGATCCCTGTCTTCAAGAACATGGCTGATGCCATGAGGAAGCACC CGGAGGTAG  ATGTGCTCATCAACTTTGCCTCTCTCCGCTCTGCCTATGACAGCACCATGGAGACCAT  GAACTATGCCCAGATCCGGACCATCGCCATCATAGCTGAAGGCATCCCTGAGGCCCTC  ACGAGAAAGCTGATCAAGAAGGCGGACCAGAAGGGAGTGACCATCATCGACCTGCCA  CTGTTGGAGGCATCAAGCCTGGGTGCTTTAAGATTGGCAACACAGGTGGGATGCTGGA  CAACATCCTGGCCTCCAAACTGTACCGCCCAGGCAGCGTGGCCTATGTCTCAGTTCC  GGAGGCATGTCCAACGAGCTCAACAATATCATCTCTCGACACCGGATGGCGTCTATG  AGGGCGTGGCCATTGGTGGGGACAGGTACCCGGGCTCCACATTGATGGATCATGTGTT  ACGCTATCAGGACACTCCAGGAGTCAAATGATTGTGGTTCTTGAGAGATTGGGGGC  ACTGAGGAATATAAGATTGCGCGGGGCATCAAGGAGGGCCGCTCCTAAGCCCATCG  TCTGCTGGTGCATCGGGACGTGTGCCACCATTGTTCTCTCTGAGGTCCAGTTTGGCCA  TGCTGGAGCTTGTGCCAACAGGCTTCTGAACTGCAGTAGCCAAGAACCAGGCTTTG  AAGGAAGCAGGAGTGTGTTGTGCCCGGAGCTTGTATGAGCTTGGAGAGATCATCCAGT  CTGTATACGAAGATCTCGTGGCCAATGGAGTCATTGTACCTGCCAGGAGGTGCCGCC  CCCAACCGTGCCCATGGACTACTCCTGGGCCAGGGAGCTTGGTTTGATCCGCAAACCT  GCCTCGTTCATGACCAGCATCTGCGATGAGCGAGGACAGGAGCTCATCTACCGGGCA  TGCCCATCACTGAGGTCTTCAAGGAAGAGATGGGCATTGGCGGGTCTCGGCCTCCT  CTGGTTCCAGAAAAGGTTGCCCTAAGTACTCTTGCCAGTTCATTGAGATGTGTCTGATG  GTGACAGCTGATCACGGGCCAGCCGTCTCTGGAGCCCACAACACCATCATTTGTGCGC  GAGCTGGGAAAGACTGGTCTCCAGCCTCACCTCGGGGCTGCTCACCATCGGGGATCG  GTTTGGGGGTGCCCTTGGATGCAGCAGCCAAGATGTTTCAAGTAAAGCCTTGGACAGTGGC  ATTATCCCATGGAGTTTGTGAACAAGATGAAGAAGGAAGGAAGCTGATCATGGGCA  TTGGTCCAGGAGTGAAGTCGATAAACAACCCAGACATGCGAGTGAGATCCTCAAGA  TTACGTCAGGCAGCACTTCCCTGCCACTCCTCTGCTCGATTATGCACTGGAAGTAGAG  AAGATTACCACCTCGAAGAAGCAAATCTTATCCTGAATGTAGATGGTCTCATCGGAG  TCGCATTGTAGACATGCTTAGAACTGTGGGTCTTTACTCGGGAGGAAGCTGATGA  ATATATTGACATTGGAGCCCTCAATGGCATCTTTGTGCTGGGAAGGAGTATGGGGTTC  ATTGGACACTATCTTGATCAGAAGAGGCTGAAGCAGGGGCTGTATCGTCTATCCGTGGG  ATGATATTTATATGTTCTTCCGGAACACATGAGCATGTAAGCGGCCGCTTTTTTCTCT  T</p>
	<p>ORF Start: at 2</p> <p>ORF Stop: TAA at 3287</p>
	<p>SEQ ID NO: 238</p> <p>1095 aa</p> <p>MW at 120201.2kD</p>
<p>NOV15g, 256420210 Protein Sequence</p>	<p>QNSTMSAKAISEQTGKELLYKFICTTSAIQNRFKYARVTPD TDWARLLQDHPWLLSQN  LVVKPDQLIKRRGKLGVLGVNLTLDGVKSWLKPRLGQEATVGKATGFLKNFLIEPFVP  HSQAEEFYVCIYATREGDYVLFFHHEGGVDVGDVDAKAQKLLVGVDEKLNPEDIKKHLL  VHAPEDKKEILASFISGLNFYEDLYFTYLEINPLVVT KDGVVYLDLA AKVDATADYI  CKVKWGDIEFPFPGREAYPEEAYIADLDAKSGASLKLTLNPKGRIWTMVAGGGASV  VYSDTICDLGGVNELANYGEYS GAPSEQTYDYAKTILSLM TREKHPDGKILIIIGSSI  ANFTNVAATFKGIVRAIRDYQG PLKEHEVTIFVRRGGPNYQ EGLRVMGEVGTGTGPIPI  HVFGTETHMTAIVGMALGHRPIPNQPPPTAAHTANFLN ASGSTSTPAPSRTASFSES  ADEVAPAKKAKPAMPQKGSTTLFSRHTKAI VWGMQTRAVQGM LDFDYVCSRDEPSVAA  MVYPFTGDHKQKFYWGHEKILIPVFKNMADAMRKHPEVDVLINFASLR SAYDSTMETM  NYAQIRTIAIIAEGIP EALTRKLIKADQKGVTTIIGPATVGGIKPGCFKIGNTGGM LD  NILASKLYRPGSVAYVSRSGMSNELNNIISRTTDGVYEGVAIGGDRYPGSTFMDHVL  RYQDTPGVKMIIVLGEIIGTTEYKICRG IKEGRLTKPIVCWIGCTCATMFSSVEQVFGH  AGACANQASETAVAKNQALKEAGVFVPRSFDELGEIIQSVYEDLVANGVIVPAQEVFP  PTVPM DYSWARELG LIRKPASFMTSICDERGQELIYAGMPITEVFKEEMGIGGV LGLL  WFQKRLPKYSCQFIEMCLMVTADHGPAVSGAHNTIICARAGKDLVSSLTSGLLTIGDR  FGGALDAAAKMFSKAFDSGII PMEFVNKMKKEGKLIMGIGHRVKSINNPDMRVQILKD</p>

	YVRQHFPATPLLDYALEVEKITTSKKPNLILNVDGLIGVAFVDMRLNCGSFRTREEADE YIDIGALNGIFVLGRSMGFI GHYLDQKRLKQGLYRHPWDDISYVLPEHMSM		
	SEQ ID NO: 239	2290 bp	
NOV15h, 256202925 DNA Sequence	CCAGAATTCCACCATGTCGGCCAAGGCAATTTTCAGAGCAGACGGGCAAAGAACTCCTT TACAAGTTTCATCTGTACCACCTCAGCCATCCAGAATCGGTTCAAGTATGCTCGGGTCA CTCCTGACACAGACTGGGCCCGCTTGCTGCAGGACCACCCCTGGCTGCTCAGCCAGAA CTGGGTAGTCAAGCCAGACCAGCTGATCAAACGTCGTGGAAACTTGGTCTCGTTGGG GTCAACCTCACTCTGGATGGGGTCAAGTCCTGGCTGAAGCCACGGCTGGGACAGGAAG CCACAGTGAGTGGGCATGGGGTCAAGATGAACGTGTGTGGTAACAGAAGCAAATATGG TCACCTTCAGGTTGGCAAGGCCACAGGCTTCCTCAAGAACTTTCTGATCGAGCCCTTC GTCCCCCACAGTCAGGCTGAGGAGTTCTATGTCTGCATCTATGCCACCCGAGAAGGGG ACTACGTCTGTTCCACCACGAGGGGGGTGTGGACGTGGGTGATGTGGACGCCAAGGC CCAGAAGCTGCTTGTGGCGTGGATGAGAACTGAATCCTGAGGACATCAAAAAACAC CTGTTGGTCCACGCCCTGAAGACAAGAAAGAAATCTGGCCAGTTTATCTCCGGCC TCTTCAATTCTACGAGGACTTGACTTCACCTACCTCGAGATCAATCCCTTGTAGT GACCAAAGATGGAGTCTATGTCCTTGACTTGGCGGCCAAGGTGGACGCCACTGCCGAC TACATCTGCAAAGTGAAGTGGGGTGACATCGAGTTCCTCCCCCTTCGGGCGGGAGG CATATCCAGAGGAAGCTACATTGCAGACCTCGACGCCAAAAGTGGGGCAAGCCTGAA GCTGACCTTGCTGAACCCCAAAGGAGGATCTGGACCATGGTGGCCGGGGGTGGCGCC TCTGTCTGTACAGCGATACCATCTGTGATCTAGGGGGTGTCAACGAGCTGGCAAAC ATGGGGAGTACTCAGCGCCCCCAGCGAGCAGCAGACCTATGACTATGCCAAGACTAT CCTCTCCCTCATGACCCGAGAGAAGCACCAGATGGCAAGATCCTCATCATTGGAGGC AGCATCGCAAACCTTCAACAACGTGGCTGCCACGTTCAAGGGCATCGTGAGAGCAATTC GAGATTACCAGGGCCCCCTGAAGGAGCACGAAGTCACAATCTTTGTCCGAAGAGGTGG CCCCACCTATCAGGAGGGCTTACGGGTGATGGGAGAAGTCGGGAAGACCACTGGGATC CCCATCCATGTCTTTGGCACAGAGACTCACATGACGGCCATGTGGGCATGGCCCTGG GCCACCGGCCCATCCCCAACAGCCACCCACAGCGGCCACACTGCAAACCTTCTCTCT CAACGCCAGCGGGAGCACATCGACGCCAGCCCCAGCAGGACAGCATCTTTTCTGAG TCCAGGGCCGATGAGGTGGCGCCTGCAAAGAAGGCCAAGCCTGCCATGCCACAAGGAA AGAGCACCAACCTCTTCAGCCGCCACCAAGGCCATTGTGTGGGGCATGCAGACCCG GGCCGTGCAAGGCATGCTGGACTTGACTATGTCTGCTCCGAGACGAGCCCTCAGTG GCTGCCATGGTCTACCCCTTCTACTGGGGACCACAAGCAGAAGTTTTACTGGGGGCACA AAGAGATCCTGATCCCTGTCTTCAAGAACATGGCTGATGCCATGAGGAAGCACCCGGA GGTAGATGTGCTCATCAACTTTGCTTCTCTCCGCTCTGCCTTGGATGCAGCAGCCAAG ATGTTCAAGTAAAGCCTTTGACAGTGGCATTATCCCCATGGAGTTTGTGAACAAGATGA AGAAGGAAGGGAAGCTGATCATGGGCATTGGTCACCGAGTGAAGTCGATAAACAAACC AGACATGCGAGTGCAGATCCTCAAAGATTACGTACGGCAGCACTTCCCTGCCACTCCT CTGCTCGATTATGCACTGGAAGTAGAGAAGATTACCACCTCGAAGAAGCCAAATCTTA TCCTGAATGTAGATGGTCTCATCGGAGTCGATTTGTAGACATGCTTAGAACTGTGG GTCCTTTACTCGGGAGGAAGCTGATGAATATATTGACATTGGAGCCCTCAATGGCATC TTTGTGCTGGGAAGGAGTATGGGGTTCATTGGACACTATCTTGATCAGAAGAGGCTGA AGCAGGGGCTGTATCGTATCCGTGGGATGATATTTATATGTTCTTCCGGAACACAT GAGCATGTAAGCGGCCGCTTTTTCTCT		
	ORF Start: at 2		ORF Stop: TAA at 2270
	SEQ ID NO: 240	756 aa	MW. at 83890.7kD
NOV15h, 256202925 Protein Sequence	QNSTMSAKAISEQTGKELLYKFICTTSAIQNRFKYARVTPDTDWARLLQDHPWLLSQN LVVKPDQLIKRRGKLGVLGVNLTLDGVKSWLKPRLGQEATVSGHGVKMNVCNRSKYG HLQVGKATGFLKNFLIEPFVPHSQAEFVFCIYATREGDYVLFHHEGGVDVGDVDAKA QKLLVGVDEKLNPEIDIKHLLVHAPEDKKEILASFISGLNFNYEDLYFTYLEINPLVV TKDGVYVLDLAAKVDATADYICKVKWGDIETPPPGREAYPEEAYIADLDAKSGASLK LTLNPKGRIWTMVAGGGASVVYSDTICDLGGVNELANYGEYSAGPSEQQTYDYAKTI LSLMTREKHPDGKILIIIGGSIANFTNVAATFKGIVRAIRDYQGPKHEHEVTIFVRRGG PNYQEGLRVMGEVGKTTGIPIHVFGTETHMTAIVGMALGHRPIPNQPPTAAHTANFLL NASGSTSTPAPSRASFSERADEVAPAKKAKPAMPQKSTTLFSRHTKAIVWGMQTR AVQGMLEDFDYCSRDEPSVAAMVYPFTGDHKQKFYWGHEILIPVFNKMDAMRKHPE VDVLINFASLRSALDAAKMFSAFDSGIIPMEFVNKMKKEGKLIMGIGHRVKSINNP DMRVRILKDYVRQHFPATPLLDYALEVEKITTSKKPNLILNVDGLIGVAFVDMRLNCG SFRTREEADEYIDIGALNGIFVLGRSMGFI GHYLDQKRLKQGLYRHPWDDISYVLPEHM		

	SM
	SEQ ID NO: 241 3310 bp
NOV15i, 259856081 DNA Sequence	<p> CACCATGTCGGCCAAGGCAATTTTCAGAGCAGACGGGCAAAGAAGCTCCTTTACAAGTTC  ATCTGTACCACTCAGCCATCCAGAATCGGTTCAAGTATGTCGGGTCACTCCTGACA  CAGACTGGGCCCCGCTTGCTGCAGGACCACCCCTGGCTGCTCAGCCAGAAGTTGGTAGT  CAAGCCAGACCAGCTGATCAAACGTCGTGGAAGAACTTGGTCTCGTTGGGGTCAACCTC  ACTCTGGATGGGGTCAAGTCTCGCTGAAGCCACGGCTGGGACAGGAAGCCACAGTTG  GCAAGGCCACAGGCTTCCTCAAGAACTTTCTGATCGAGCCCTTCGTCCCCACAGTCA  GGCTGAGGAGTTCTATGTCTGCATCTATGCCACCCGAGAAGGGGACTACGTCCTGTTC  CACCACGAGGGGGGTGTGGACGTGGGTGATGTGGACGCCAAGGCCAGAGCTGCTTG  TTGGCGTGGATGAGAACTGAATCCTGAGGACATCAAAAAACACCTGTTGGTCCACGC  CCCTGAAGACAAGAAAGAAATTTGGCCAGTTTATCTCCGGCCTCTTCAATTTCTAC  GAGGACTTGTAATTCACCTACCTCGAGATCAATCCCTTGTAGTGACCAAAGATGGAG  TCTATGTCCTTGACTGGCGGCCAAGGTGGACGCCACTGCCGACTACATCTGCAAAGT  GAAGTGGGGTGACATCGAGTTCCCTCCCCCTTCGGGCGGGAGGCATATCCAGAGGAA  GCCTACATTGCAGACCTCGATGCCAAAGTGGGGCAAGCCTGAAGCTGACCTTGCTGA  ACCCCAAAGGGAGGATCTGGACCATGGTGGCCGGGGGTGGCGCCTCTGTCGTGTACAG  CGATACCATCTGTGATCTAGGGGGTGTCAACGAGCTGGCAAAGTATGGGGAGTACTCA  GGCGCCCCAGCGAGCAGCAGACCTATGATTATGCCAAGACTATCCTCTCCCTCATGA  CCCGAGAGAAGCACCAGATGGCAAGATCCTCATCATTGGAGGCAGCATCGCAAACCT  CACCAACGTGGCTGCCACGTTCAAGGGCATCGTGAGAGCAATTCGAGATTACCAGGGC  CCCCGAAGGAGCACGAAGTCACAATCTTTGTCCGAAGAGGTGGCCCCAAGTATCAGG  AGGGCTTACGGGTGATGGGAGAAGTCGGGAAGACCACTGGGATCCCCATCCATGTCTT  TGGCACAGAGACTCACATGACGGCCATTGTGGGCATGGCCCTGGGGCACCGGCCCATC  CCCAACGAGCCACCCACAGCGGCCACACTGCAAACCTCCTCAACGCCAGCGGGA  GCACATCGACGCCAGCCCCAGCAGGACAGCATCTTTTCTGAGTCCAGGCCGATGA  GGTGGCGCCTGCAAAGAAGGCCAAGCCTGCCATGCCACAAGATTAGTCCCAAGTCCA  AGATCCCTGCAAGGAAAGAGCACCACCTCTCAGCCGCCACACCAAGGCCATTGTGT  GGGGCATGCAGACCCGGGCCGTGCAAGGCATGCTGGACTTTGACTATGTCTGCTCCCG  AGACGAGCCCTCAGTGGCTGCCATGGTCTACCTTTCACTGGGGACCACAAGCAGAAG  TTTTACTGGGGGCACAAAGAGATCCTGATCCCTGTCTTCAAGAACATGGCTGATGCCA  TGAGGAAGCACCCGAGGTAGATGTGCTCATCAACTTTGCCTCTCTCCGCTCTGCCTA  TGACAGCACCATGGAGACCATGAAGTATGCCAGATCCGACCATCGCCATCATAGCT  GAAGGCATCCCTGAGGCCCTCACGAGAAAGCTGATCAAGAAGGCGGACCAGAAGGGAG  TGACCATCATCGGACCTGCCACTGTTGGAGGCATCAAGCCTGGGTGCTTTAAGATTGG  CAACACAGGTGGGATGCTGGACAACATCCTGGCCTCCAAACTGTACCGCCAGGCAGC  GTGGCCTATGTCTCACGTTCCGGAGGCATGTCCAACGAGCTCAACAATATCATCTCTC  GGACCACGGATGGCGTCTATGAGGGCGTGGCCATTGGTGGGGACAGGTACCCGGGCTC  CACATTCATGGATCATGTGTACGCTATCAGGACACTCCAGGAGTCAAAATGATTGTG  GTTCTTGGAGAGATTGGGGGCACTGAGGAATATAAGATTGCGGGGCATCAAGGAGG  GCCGCCTCACTAAGCCCATCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT  CTCTGAGGTCCAGTTTGGCCATGCTGGAGCTTGTGCCAACCAGGCTTCTGAAACTGCA  GTAGCCAAGAACCAGGCTTTGAAGGAAGCAGGAGTGTGTGCCCCGAGCTTTGATG  AGCTTGAGAGATCATCCAGTCTGTATACGAAGATCTCGTGGCCAAATGGAGTCATTGT  ACCTGCCCAGGAGGTGCCGCCCCAACCGTGCCCATGGACTACTCCTGGGCCAGGGAG  CTTGGTTTGATCCGCAAACCTGCCTCGTTCATGACCAGCATCTGCGATGAGCGAGGAC  AGGAGCTCATCTACGCGGGCATGCCCATCACTGAGGTCTTCAAGGAAGAGATGGGCAT  TGGCGGGGTCTCGGCCTCCTTGGTTCCAGAAAGGTTGCCTAAGTACTCTTGCCAG  TTCATTGAGATGTGTCTGATGGTGACAGCTGATCACGGGCCAGCCGTCTCTGGAGCCC  ACAACACCATCATTTGTGCGCGAGCTGGGAAAGACCTGGTCTCCAGCCTCACCTCGGG  GCTGCTCACCATCGGGGATCGGTTTGGGGGTGCCTTGATGCAGCAGCCAAGATGTTT  AGTAAAGCCTTTGACAGTGGCATTATCCCATGGAGTTTGTGAACAAGATGAAGAAG  AAGGGAAGCTGATCATGGGCATTGGTCACCGAGTGAAGTCGATAAACAACCCAGACAT  GCGAGTGCAGATCCTCAAAGATTACGTCAAGCAGCACTTCCCTGCCACTCCTCTGCTC  GATTATGCACTGGAAGTAGAGAAGATTACACCTCGAAGAAGCCAATCTTATCCTGA  ATGTAGATGGTCTCATCGGAGTCGCAATTTGTAGACATGCTTAGAACTGTGGGTCCCT  TACTCGGGAGGAAGCTGATGAATATATTGACATTGGAGCCCTCAATGGCATCTTTGTG  CTGGGAAGGAGTATGGGGTTCATTGGACACTATCTTGATCAGAAGAGGCTGAAGCAGG  GGCTGTATCGTCATCCGTGGGATGATATTTATATGTTCTTCCGGAACATGAGCAT  GTAA </p>

	ORF Start: at 2		ORF Stop: TAA at 3308
	SEQ ID NO: 242	1102 aa	MW at 120939.0kD
NOV15i, 259856081 Protein Sequence	<p>TMSAKAISEQTGKELLYKFICTTSAIQNRFKYARVTPDTPDWARLLQDHPWLLSQNLVVKPDQLIKRRGKLGVLGVNLTLDGVKSWLKPRLGQEATVGKATGFLKNFLIEPFPVPHSQAEFFYVCIYATREGDYVLFHHEGGVDVGDVDAKAQKLLVGVEKLNPEDIKKHLLVHAPEDKKEILASFISGLNFYEDLYFTYLEINPLVVTKDGVYVLDLAALKVDATADYICKVKWGDIEFPPFPFGREAYPEEAYIADLDAKSGASLKLTLNPKGRIWTMVAGGGASVVYSDTICDLGGVNELANYGEYS GAPSEQQTYDYAKTILSLMTREKHPDGKILIIIGGSIANFTNVAATFKGIVRAIRDYQGPLEHEVTIFVRRGGPNYQEGLRVMGEVGTGTGPIPIHVGTTETHMTAIVGMALGHRPIPNQPPPTAHTANFLNAGSTSTPAPSRITASFSERADEVAPAKKAKPAMPQDSVPSRSLQKSTTLFSRHTKAI VWGMQTRAVQGMDFDYVCSRDEPSVAAMVYPFTGDHKQKFYWGHEKILIPVFKNMADAMRKHPEVDVLINFASLSAYDSTMETMNYAQIRTIATIAEGIPEALTRKLIKADQKGVTIIGPATVGGIKPGCFKIGNTGGMLDNILASKLYRPGSVAYVSRSGGMSNELNNIISRTTDGVYEGVAIGGDRYPGSTFMDHVLRYQDTPGVKMIIVLGEIGGTEEYKICRGIKEGRLTKPIVCWCIGTCATMFSSEVQFGHAGACANQASETAVAKNQALKEAGVFVPRSFDELGEIIQSVYEDLVANGVIVPAQEVPPPTVPMDYSWARELGLIRKPASFMSTICDERGQELIYAGMPITEVFKEEMGIGGVLLGLWFQKRLPKYSCQFIEMCLMVTADHGPAVSGAHTTICARAGKDLVSSLTSGLLTIGDRFGGALDAAKMFSKAFDSGII PMEFVNKMKKEGKLIMGIGHRVKSINNPD MRVQILKDYVRQHFPATPLLDYALEVEKITTSKKNLILNVDGLIGVAFVDMLRNCGSFTREEADEYIDIGALNGIFVLGRSMGFIGHYLDQKRLKQGLYRHPWDDISYVLPEHMSM</p>		
	SEQ ID NO: 243	3317 bp	
NOV15j, 256388552 DNA Sequence	<p>CCAGAATTCACCATGTGCGCCAAGGCAATTTAGAGCAGACGGGCAAGAAGCTCCTTTACAAGTTCATCTGTACCACCTCAGCCATCCAGAATCGGTTCAAGTATGCTCGGGTCACTCCTGACACAGACTGGGCCCGCTTGCTGCAGGACCACCCCTGGCTCTCAGCCAGAACTTGCTAGTCAAGCCAGACCAGCTGATCAAACGTCGTGAAAACCTTGGTCTCGTTGGGGTCAACCTCACTCTGGATGGGGTCAAGTCTGGCTGAAGCCACGGCTGGGACAGGAAGCCACAGTTGGCAAGGCCACAGGCTTCCCTCAAGAAGCTTTCTGATCGAGCCCTTCGTCCTCCACAGTCAGGCTGAGGAGTTCTATGTCTGCATCTATGCCACCCGAGAAGGGGACTACGTCTGTCTCCACCACGAGGGGGGTGTGGACGTGGGTGATGTGGACGCCAAGGCCAGAGCTGCTTGTGGCGTGGATGAGAACTGAATCCTGAGGACATCAAAAAACCTGTTGGTCCACGCCCCCTGAAGACAAGAAAGAAATTTCTGGCCAGTTTTATCTCCGGCCTCTTCAATTTCTACGAGGACTTGTACTTCACCTACCTCGAGATCAATCCCTTGTAGTGACCAAGATGGAGTCTATGTCCTTGACTTGGCGGCCAAGGTGGACGCCACTGCCGACTACATCTGCAAAGTGAAGTGGGGTGACATCGAGTTCCTTCCCCCTTCGGGCGGGAGGCATATCCAGAGGAAGCCTACATTGTCAGACCTCGATGCCAAAAGTGGGGCAAGCCTGAAGCTGACCTTGCTGAACCCCAAGGGAGGATCTGGACCATGGTGGCGGGGGTGGCGCCTCTGTCTGTGTACAGCGATACCATCTGTGATCTAGGGGGTGTCAACGAGCTGGCAAACATATGGGGAGTACTCAGGCGCCCCCAGCGAGCAGCAGACCTATGATTATGCCAAGACTATCCTCTCCCTCATGACCCGAGAGAAGCAGCCAGATGGCAAGATCCTCATATTGGAGGCAGCATCGCAAACCTTACCAACGTCGCTGCCACGTTCAAGGGCATCGTGAGAGCAATTCGAGATTACCAGGGCCCCCTGAAGGAGCACGAAGTCACAATCTTTGTCCGAAGAGGTGGCCCCA ACTATCAGGAGGGCTTACGGGTGATGGGAGAAGTCGGGAAGACCACTGGGATCCCCATCCATGTCTTTGGCACAGAGACTCACATGACGGCCATTGTGGGCATGGCCCTGGGCCACCGGCCCATCCCCAACAGCCACCCACAGCGGCCACACTGCAAACCTTCTCTCAACGCCAGCGGGAGCATCGACGCCAGCCCCAGCAGGACAGCATCTTTTCTGAGTCCAGGGCGATGAGGTGGCGCCTGCAAAGAAGGCCAAGCCTGCCATGCCACAAGATTGAGTCCAAAGTCCAAGATCCCTGCAAGGAAAGAGCACCACCTCTTCAGCCGCCACACCAAGGCCATTTGTGTGGGGCATGCAGACCCGGGCGGTGCAAGGCATGCTGGACTTTGACTATGTCTGTCTCCCGAGACGAGCCCTCAGTGGCTGCCATGGTCTACCTTTCACTGGGGACCACAAGCAGAAAGTTTTACTGGGGGCACAAAGAGATCCTGATCCCTGTCTTCAAGAACATGCTGATGCCATGAGGAAGCACCCGGAGGTAGATGTGCTCATCAACTTTGCTCTCTCCGCTCTGCCTATGACAGCACCATTGGAGACCATGAACATATGCCAGATCCGGACCATCGCCATCATAGCTGAAGGCATCCCTGAGGCCCTCACGAGAAAGCTGATCAAGAAGGCGGACAGAGGGAGTGACCATCATCGGACCTGCCACTGTTGGAGGCATCAAGCCTGGGTGCTTTAAGATTGGCAACACAGGTGGGATGCTGGACAACATCCTGGCCTCCAAACTGTACCGCCAGGCAGCGTGGCCTATGTCTCACGTTCCGGAGGCATGTCCAACGAGCTCAACAATATCATCTCTCGGACCAAGGATGGCGTCTATGAGGGCGTGGCCATTGTTGGGACAGGTATACCGGGCTCCACATTGATGGATCATGTGTTACGCTATCAGGACACTCCAGGAGTCAAA</p>		

	<p>ATGATTGTGGTTCTTGGAGAGATTGGGGGCACTGAGGAATATAAGATTGCGGGGCA  TCAAGGAGGGCCGCTCACTAAGCCCATCGTCTGCTGGTGACATCGGGACGTGTGCCAC  CATGTTCTCCTCTGAGGTCCAGTTTGGCCATGCTGGAGCTTGTGCCAACCAGGCTTCT  GAAACTGCAGTAGCCAAGAACCAGGCTTTGAAGGAAGCAGGAGTGTGTTGCCCCGGA  GCTTTGATGAGCTTGGAGAGATCATCCAGTCTGTATACGAAGATCTCGTGGCCAATGG  AGTCATTGTACCTGCCCAGGAGGTGCCGCCCCAACCGTGCCCATGGAATACTCCTGG  GCCAGGGAGCTTGGTTTGTATCCGCAAACTGCCTCGTTTCATGACCAGCATCTGCGATG  AGCGAGGACAGGAGCTCATCTACGCGGGCATGCCATCACTGAGGTCTTCAAGGAAGA  GATGGGCATTGGCGGGTCTCGCCTCCTCTGGTTCAGAAAAGTTGCCCTAAGTAC  TCTTGCCAGTTTATTGAGATGTGTCTGATGGTGACAGCTGATCACGGGCCAGCCGTCT  CTGGAGCCCAACAACCATCATTGTGCGCGAGCTGGGAAAGACCTGGTCTCCAGCCT  CACCTCGGGGCTGCTACCATCGGGGATCGGTTTGGGGGTGCTTGGATGCAGCAGCC  AAGATGTTTCAGTAAAGCCTTTGACAGTGGCATTATCCCCATGGAGTTTGTGAACAAGA  TGAAGAAGGAAGGGAAGCTGATCATGGGCATTGGTCACCGAGTGAAGTCGATAAACA  CCCAGACATGCGAGTGCAGATCCTCAAAGATTACGTCAGGCAGCACTTCCCTGCCACT  CCTCTGCTCGATTATGCACTGGAAGTAGAGAAGATTACCACCTCGAAGAAGCCAAATC  TTATCCTGAATGTAGATGGTCTCATCGGAGTGCATTGTAGACATGCTAGAAACTG  TGGGTCTTTACTCGGGAGGAAGCTGATGAATATATTGACATTGGAGCCCTCAATGGC  ATCTTTGTGCTGGGAAGGAGTATGGGGTTCATTGGACACTATCTTGATCAGAAGAGGC  TGAAGCAGGGGCTGTATCGTATCCGTGGGATGATATTCATATGTTCTTCCGGAACA  CATGAGCATGT</p>
	<p>ORF Start: at 2</p> <p>ORF Stop: end of sequence</p>
	<p>SEQ ID NO: 244</p> <p>1106 aa</p> <p>MW at 121268.4kD</p>
<p>NOV15j, 256388552 Protein Sequence</p>	<p>QNSTMSAKAISEQTGKELLYKFICTTSAIQNRFKYARVTPDWDWARLLQDHPWLLSQN  LVVKPDQLIKRRGKLGVLGVNLTLDGVKSWLKPRLGQEATVGKATGFLKNFLIEFPVP  HSQAEEFYVCIYATREGDYVLFHHEGGVDVGDVDAKAQKLLVGVDEKLNPEDIKKHLL  VHAPEDKKEILASFISGLNFYEDLYFTYLEINPLVVTKDGVVYLDLAAKVDATADYI  CKVKWGDIEFPFPFGREAYPEEAYIADLDAKSGASLKLTLNPKGRIWTMVAGGGASV  VYSDTICDLGGVNELANYGEYSAPSEQQTYDYAKTILSLMTREKHPDGKILIIIGGSI  ANFTNVAATFKGIVRAIRDYQGPLKEHEVTIFVRRGGPNYQEGLRVMGEVGGKTTGIPI  HVFGTETHMTAIVGMALGHRPIPNQPPTAAHTANFLNLSGSTSTPAPSRTASFSES  ADEVAPAKKAKPAMPQDSVPSRSLQKSTTLFSRHTKAIWGMQTRAVQGMDFDYV  CSRDEPSVAAMVYPFTGDHKQKPYWGHKEILIPVFKNMADAMRKHPVDVLINFASLR  SAYDSTMETMNYAQIRTIATIAEGIPEALTRKLIKADQKGVTIIGPATVGGIKPGCF  KIGNTGMLDNILASKLYRPGSVAYVSRSGMSNELNNIISRTTDGVYEGVAIGGDRY  PGSTFMDHVLRYQDTPGVKMIIVLGEIGGTEEYKICRGIKEGRLTKPIVCWCIGTCAT  MFSSEVQFGHAGACANQASETAVAKNQALKEAGVFVPRSFDELGEIIQSVEYEDLVANG  VIVPAQEVPPTVPMDSWARELGLIRKPASFMSTICDERGQELIYAGMPITEVFKEE  MGIGGVLLGLLWFQKRLPKYSCQFIEMCLMVTADHGPVSGAHTNIIICARAGKDLVSSL  TSGLLTIGDRFGGALDAAAKMFSKAFDSGIIPMEFVNKMKKEGKLIMGIGHRVKSINN  PDMRVQILKDYVRQHPATPLLDYALEVEKITTSKKPNLILNVDGLIGVAFVDMRLNC  GSFTREEADEYIDIGALNGIFVLGRSMGFIGHYLDQKRLKQGLYRHPWDDISYVLEPH  MSMX</p>
	<p>SEQ ID NO: 245</p> <p>3307 bp</p>
<p>NOV15k, 256420210 DNA Sequence</p>	<p>CCAGAATTCACCATGTGCGCCAAGGCAATTTTCAGAGCAGACGGGCAAAGAATCCTT  TACAAGTTCATCTGTACCACCTCAGCCATCCAGAATCGGTTCAAGTATGCTCGGGTCA  CTCCTGACACAGACTGGGCCGCTTGTGTCAGGACCACCCCTGGCTGCTCAGCCAGAA  CTTGGTAGTCAAGCCAGACCAGCTGATCAAACGTCGTGGAAAACTTGGTCTCGTTGGG  GTCAACCTCACTCTGGATGGGGTCAAGTCTGGCTGAAGCCACGGCTGGGACAGGAAG  CCACAGTTGGCAAGGCCACAGGCTTCTCAAGAATTTCTGATCGAGCCCTTCGTCCC  CCACAGTCAGGCTGAGGAGTTCTATGTCTGCATCTATGCCACCCGAGAAGGGGACTAC  GTCCTGTTCCACCACGAGGGGGGTGTGGACGTGGGTGATGTGGACGCCAAGGCCAGAG  AGCTGCTTGTGGCGTGGATGAGAACTGAATCCTGAGGACATCAAAAAACACCTGTT  GGTCCACGCCCCGTAAGACAAGAAAGAAATTCTGGCCAGTTTATCTCCGGCCTCTTC  AATTTCTACGAGGACTTGTACTTCACCTACCTCGAGATCAATCCCCTTGTAGTGACCA  AAGATGGAGTCTATGTCTTGAATTGGCGGCCAAGGTGGACGCCACTGCCGACTACAT  CTGCAAAGTGAAGTGGGGTGACATCGAGTTCCCTCCCCCTTCGGGCGGGAGGCATAT</p>

<p>CCAGAGGAAGCCTACATTGCAGACCTCGATGCCAAAAGTGGGGCAAGCCTGAAGCTGA  CCTTGCTGAACCCCAAAGGGAGGATCTGGACCATGGTGGCCGGGGTGGCGCCTCTGT  CGTGATACAGCGATACCATCTGTGATCTAGGGGGTGTCAACGAGCTGGCAAACCTATGGG  GAGTACTCAGGCGCCCCAGCGAGCAGCAGACCTATGACTATGCCAAGACTATCCTCT  CCCTCATGACCCGAGAGAAGCACCCAGATGGCAAGATCCTCATCATTGGAGGCAGCAT  CGCAAACCTTACCAACGTGGCTGCCACGTTCAAGGGCATCGTGAGAGCAATTCGAGAT  TACCAGGGCCCCCTGAAGGAGCACGAAGTCACAATCTTTGTCCGAAGAGGTGGCCCCA  ACTATCAGGAGGGCTTACGGGTGATGGGAGAAGTCGGGAAGACCACTGGGATCCCCAT  CCATGTCTTTGGCACAGAGACTCACATGACGGCCATTGTGGGCATGGCCCTGGGCCAC  CGGCCCCATCCCCAACAGCCACCCACAGCGGCCCACTGCAAACTTCTCTCTCAACG  CCAGCGGGAGCACATCGACGCCAGCCCCAGCAGGACAGCATCTTTTCTGAGTCCAG  GGCCGATGAGGTGGCGCCTGCAAAGAAGGCCAAGCCTGCCATGCCACAAGGAAAGAGC  ACCACCTCTTCAGCCGCCACACCAAGGCCATTGTGTGGGGCATGCAGACCCGGGCCG  TGCAAGGCATGCTGGACTTTGACTATGTCTGCTCCCGAGACGAGCCCTCAGTGGCTGC  CATGGTCTACCTTTCACTGGGGACCACAAGCAGAAGTTTACTGGGGGCACAAAGAG  ATCCTGATCCCTGTCTTCAAGAACATGGCTGATGCCATGAGGAAGCACCCGGAGGTAG  ATGTGCTCATCAACTTTCCTCTCTCCGCTCTGCCTATGACAGCACCATTGGAGACCAT  GAACTATGCCCAGATCCGGACCATCGCCATCATAGCTGAAGGCATCCCTGAGGCCCTC  ACGAGAAAGCTGATCAAGAAGGCGGACAGAAAGGGAGTGACCATCATCGACCTGCCA  CTGTTGGAGGCATCAAGCCTGGGTGCTTTAAGATTGGCAACACAGGTGGGATGCTGGA  CAACATCCTGGCTCCAACTGTACCGCCCAGGCAGCGTGGCCTATGTCTCACGTTCC  GGAGGCATGTCCAACGAGCTCAACAATATCATCTCTCGGACCACGGATGGCTCTATG  AGGGCGTGGCCATTGGTGGGGACAGGTACCCGGGCTCCACATTTCATGGATCATGTGTT  ACGCTATCAGGACACTCCAGGAGTCAAAATGATTGTGGTTCTTGGAGAGATTGGGGGC  ACTGAGGAATATAAGATTGTCGGGGCATCAAGGAGGGCCGCCCTCACTAAGCCCATCG  TCTGCTGGTGCATCGGGACGTGTGCCACCATGTTCTCTCTGAGGTCCAGTTTGGCCA  TGCTGGAGCTTTGTGCCAACCAAGGCTTCTGAAACTGCAGTAGCCAAGAACCAAGCCTTG  AAGGAAGCAGGAGTGTGTTGTGCCCCGAGCTTTGATGAGCTTGGAGAGATCATCCAGT  CTGTATACGAAGATCTCGTGGCCAATGGAGTCATTGTACCTGCCCAGGAGGTGCCGCC  CCCAACCGTGCCCATGGACTACTCCTGGGCCAGGGAGCTTGGTTTGATCCGCAAACCT  GCCTCGTTCATGACCAGCATCTGCGATGAGCGAGGACAGGAGCTCATCTACGCGGGCA  TGCCCATCACTGAGGTCTTCAAGGAAGAGATGGGCATTGGCGGGGTCCTCGGCCTCTCT  CTGGTTCAGAAAAGGTTGCCCTAAGTACTCTTGCCAGTTCATTGAGATGTGTCTGATG  GTGACAGCTGATCACGGGCCAGCCGTCTCTGGAGCCCACAACACCATCATTTGTGCGC  GAGCTGGGAAAGACTGGTCTCCAGCCTCACCTCGGGGCTGCTCACCATCGGGGATCG  GTTTGGGGGTGCCCTGGATGCAGCAGCCAAGATGTTTCAGTAAAGCCTTTGACAGTGGC  ATTATCCCCATGGAGTTTGTGAACAAGATGAAGAAGGAAGGCAAGCTGATCGGGCA  TTGGTCACCGAGTGAAGTCGATAAAACAACCCAGACATGCGAGTGCAGATCCTCAAAGA  TTACGTCAGGCAGCACTTCCCTGCCACTCCTCTGCTCGATTATGCACTGGAAGTAGAG  AAGATTACCACCTCGAAGAAGCCAAATCTTATCCTGAATGTAGATGGTCTCATCGGAG  TCGCATTTGTAGACATGCTTAGAAAAGTGTGGGTCTTTACTCGGGAGGAAGCTGATGA  ATATATTGACATTGGAGCCCTCAATGGCATCTTTGTGCTGGGAAGGAGTATGGGGTTC  ATTGGACACTATCTTGATCAGAAGAGGCTGAAGCAGGGGCTGATATCGTTCATCCGTGGG  ATGATATTTATATGTTCTTCCGGAACACATGAGCATGTAAGCGGCCGCTTTTTTCTCT  T</p>			
ORF Start: at 2		ORF Stop: TAA at 3287	
SEQ ID NO: 246		1095 aa	MW at 120201.2kD
NOV15k, 256420210 Protein Sequence	<p>QNSTMSAKAISEQTGKELLYKFICTTSAIQNRFKYARVTPDPTDWARLLQDHPWLLSQN  LVVKPDQLIKRRGKLGVLGVNLTLDGVKSWLKPRLGQEAATVGKATGFLKNFLIEPFPV  HSQAEEFYVCIYATREGDYVLFHHEGGVDVGDVDAKAQKLLVGVDEKLNPEDIKKHLL  VHAPEDKKEILASFISGLNFYEDLYFTYLEINPLVVTKDGVYVLDLAAKV DATADYI  CKVKWGDIEFPFPFGREAYPEEAYIADLDAKSGASLKLTLNPKGRIWTMVAGGASV  VYSDTICDLGGVNELANYGEYS GAPSEQQTYDYAKTILSLMTREKHPDGKILIIIGGSI  ANFTNVAATFKGI VRAIRDYQGPLKEHEVTIFVRRGGPNYQEGLRVMGEVKGTTGIPI  HVFGETHTMTAIVGMALGHRPIPNQPPTAHTANFLLNAGSSTSTPAPSRTASFSES  ADEVAPAKKAKPAMPQGSTTLFSRHTKAI VWGMQTRAVQGM LDFDYVCSRDEPSVAA  MVYPFTGDHKQKPYWGHKEILIPVFKNMADAMRKHPEVDVLINFASLRSAYDSTMETM  NYAQIRTIAIIAEGIPALTRKLIKADQKGVTIIGPATVGGIKPGCFKIGNTGGM L D  NILASKLYRPGSVAYVSRSGMSNELNNIISRTTDGVYEGVAIGGD RYPGSTFMDHVL</p>		

	RYQDTPGVKMIIVVLGEIGGTEEYKICRGIKEGRLTKPIVCWCIGTCATMFSSEVQFGH AGACANQASETAVAKNQALKEAGVFVPRSFDELGEIIQSVYEDLVANGVIVPAQEVPP PTVPMDSWARELGLIRKPASFMTSICDERGQELIYAGMPITEVFKEEMGIGGVLGLL WFQKRLPKYSCQFIEMCLMVTADHGPVSGAHNTIICARAGKDLVSSLTSGLLTIGDR FGGALDAAAKMFSKAFDSGIIPMEFVNKMKKEGKLIMGIGHRVKSINNPDMRVQILKD YVRQHFPATPLLDYALEVEKITTSKKPNLIILNVDGLIGVAFVMDLRNCGSFTREEADE YIDIGALNGIFVLGRSMGFIGHYLDQKRLKQGLYRHPWDDISYVLPHEHMSM		
	SEQ ID NO: 247	2290 bp	
NOV151, 256202925 DNA Sequence	CCAGAATTCCACCATGTCGGCCAAGGCAATTTAGAGCAGACGGGCAAAGAACTCCTT TACAAGTTCATCTGTACCACCTCAGCCATCCAGAATCGGTTCAAGTATGCTCGGGTCA CTCCTGACACAGACTGGGCCCCGCTTGCTGCAGGACCACCCCTGGCTGCTCAGCCAGAA CTTGGTAGTCAAGCCAGACCAGCTGATCAAACGTCGTGGAAACTTGGTCTCGTTGGG GTCAACCTCACTCTGGATGGGGTCAAGTCTCTGGCTGAAGCCACGGCTGGGACAGGAAG CCACAGTGAGTGGGCATGGGGTCAAGATGAACGTGTGTGGTAACAGAAGCAAATATGG TCACCTTCAGGTTGGCAAGGCCACAGGCTTCCTCAAGAACTTTCTGATCGAGCCCTTC GTCCCCCACAGTCAGGCTGAGGAGTTCTATGTCTGCATCTATGCCACCCGAGAAGGGG ACTACGTCTGTTCACCACGAGGGGGGTGTGGACGTGGGTGATGTGGACGCCAAGGC CCAGAAGCTGCTTGTGGCGTGGATGAGAACTGAATCCTGAGGACATCAAAAAACAC CTGTTGGTCCACGCCCTGAAGACAAGAAAGAAATTTCTGGCCAGTTTATCTCCGGCC TCTTCAATTCTACGAGGACTTGTACTTCACCTACCTCGAGATCAATCCCTTGTAGT GACCAAGATGGAGTCTATGTCTTGAATTTGGCGGCCAAGGTGGACGCCACTGCCGAC TACATCTGCAAAGTGAAGTGGGGTGACATCGAGTTCCTCCCCCTTCGGGCGGGAGG CATATCCAGAGGAAGCCTACATTGCAGACCTCGACGCCAAAAGTGGGGCAAGCCTGAA GCTGACCTTGCTGAACCCCAAAGGGAGGATCTGGACCATGGTGGCCGGGGGTGGCGCC TCTGTCTGTACAGCGATACCATCTGTGATCTAGGGGGTGTCAACGAGCTGGCAAAC ATGGGGAGTACTCAGGCGCCCCCAGCGAGCAGCAGACCTATGACTATGCCAAGACTAT CCTCTCCCTCATGACCCGAGAGAAGCACCAGATGGCAAGATCCTCATCATTTGGAGGC AGCATCGAAACTTCACCAACGTGGCTGCCACGTTCAAGGGCATCGTGAGAGCAATTC GAGATTACCAGGGCCCCCTGAAGGAGCAGAACTCACAATCTTTGTCCGAAGAGGTGG CCCCAACTATCAGGAGGGCTTACGGGTGATGGGAGAAGTCGGGAAGACCACTGGGATC CCCATCCATGTCTTGGCACAGAGACTCACATGACGGCCATTGTGGGCATGGCCCTGG GCCACCGGCCCATCCCCAACAGCCACCCACAGCGGCCACACTGCAAACCTTCTCCT CAACGCCAGCGGGAGCACATCGACGCCAGCCCCAGCAGGACAGCATCTTTTCTGAG TCCAGGGCCGATGAGGTGGCGCCTGCAAAGAAGGCCAAGCCTGCCATGCCACAAGGAA AGAGCACCACCTCTCAGCCGCCACACCAAGGCCATTGTGTGGGGCATGCAGACCCG GGCCGTGCAAGGCATGCTGGACTTGAATATGTCTGCTCCGAGACGAGCCCTCAGTG GCTGCCATGGTCTACCTTTTCACTGGGGACCAAGCAGAAAGTTTACTGGGGGCACA AAGAGATCCTGATCCCTGTCTTCAAGAACATGGCTGATGCCATGAGGAAGCACC CGGA GGTAGATGTGCTCATCAACTTTGCTTCTCTCGCTCTGCCTTGGATGCAGCAGCCAG ATGTTCAAGTAAAGCCTTTGACAGTGGCATTATCCCCATGGAGTTTGTGAACAAGATGA AGAAGGAAGGAAGCTGATCATGGGCATTGGTCACCGAGTGAAGTCGATAAACAACCC AGACATGCGAGTGCAGATCCTCAAAGATTACGTCAGGCAGCACTTCCCTGCCACTCCT CTGCTCGATTATGCACTGGAAGTAGAGAAGATTACCACCTCGAAGAAGCCAAATCTTA TCCTGAATGTAGATGGTCTCATCGGAGTCGATTTGTAGACATGCTTAGAACTGTGG GTCCTTTACTCGGGAGGAAGCTGATGAATATATTGACATTGGAGCCCTCAATGGCATC TTTGTGCTGGGAAGGAGTATGGGGTTCATTGGACACTATCTTGATCAGAAGAGGCTGA AGCAGGGGCTGTATCGTCATCCGTGGGATGATATTTTCATATGTTCTCCGGAACACAT GAGCATGTAAGCGGCCGCTTTTTCCTT		
	ORF Start: at 2		ORF Stop: TAA at 2270
	SEQ ID NO: 248	756 aa	MW at 83890.7kD
NOV151, 256202925 Protein Sequence	QNSTMSAKAISEQTGKELLYKFICTTSAIQNRFKYARVTPDTDWARLLQDHPWLLSQN LVVKPDQLIKRRKLGVLGVNLTLDGVKSWLKPRLGQEAIVSGHGVKMNVCNRSKYG HLQVGKATGFLKNFLIEPFPVPHSQAEFFVCIYATREGDYVLFHHEGGVDVGDVAKA QKLLVGVDEKLPEDIKHLLVHAPEDKKEILASFISGLNFYEDLYFTYLEINPLVV TKDGVVYVLDLAAKVDATADYICKVKWGDIEFPFPFGREAYPEEAYIADLDAKSGASLK LTLLNPKGRIWTMVGAGGASVVYSDTICDLGGVNELANYGEYS GAPSEQQTYDYAKTI LSLMTREKHPDGKILIIIGGSIANFTNVAATFKGIVRAIRDYQGPLKEHEVTIFVRRGG PNYQEGLRVMGEVGGKTTGPIHVFGTETHMTAIVGMALGHRPIPNQPPTAHTANFLL		

	NASGSTSTPAPSRTASFSESRADEVAPAKKAKPAMPQGKSTTLFSRHTKAIWGMQTR AVQGM LDFDYVCSRDEPSVAAMVYPFTGDHKQKQFYWGHEKILIPVFNK MADAMRKHPE VDVLIN FASLR SALDAAAKMFSKAFDSGII PMEFVNKMKKEGKLIMIGIHRVKSINNP DMRVRI LKDYVRQHFPATPLLDYALEVEKITTSKKPNLILNVDGLIGVAFVDMRLNCG SFTREEADEYIDIGALNGIFVLGRSMGFIGHYLDQKRLKQGLYRHPWDDISYVLP EHM SM		
	SEQ ID NO: 249	3368 bp	
NOV15m, 296463359 DNA Sequence	CCCGGTCCGAAGCGCGCGGATTCCACCATGTCGGCCAAGGCAATTCAGAGCAGACGG GCAAAGAACTCCTTTACAAGTTCATCTGTACCACCTCAGCCATCCAGAATCGGTTCAA GTATGCTCGGGTCACTCCTGACACAGACTGGGCCCGCTTGCTGCAGGACCAACCCCTGG CTGCTCAGCCAGAACTTGGTAGTCAAGCCAGACCAGCTGATCAAACGTCGTGGAACAA TTGGTCTCGTTGGGGTCAACCTCACTCTGGATGGGGTCAAGTCTGGCTGAAGCCACG GCTGGGACAGGAAGCCACAGTTGGCAAGGCCACAGGCTTCCTCAAGAACTTTCTGATC GAGCCCTTCGTCCCCACAGTCAGGCTGAGGAGTTCATATGCTGCATCTATGCCACCC GAGAAGGGGACTACGTCTCTGTTCCACCACGAGGGGGGTGTGGACGTGGGTGATGTGGA CGCCAAGGCCCAGAAGCTGCTTGTGGCGTGGATGAGAACTGAATCCTGAGGACATC AAAAAACACCTGTTGGTCCACGCCCCGTAAGACAAGAAAGAAATCTGGCCAGTTTAA TCTCCGGCCTCTTCAATTTCTACGAGGACTTGTACTTACCTACCTCAGATCAATCC CCTTGTAGTGACCAAAGATGGAGTCTATGTCTTGAAGTGGCGGCCAAGGTGGACGCC ACTGCCGACTACATCTGCAAAGTGAAGTGGGGTGCATCGAGTTCCCTCCCCCTTCG GGCGGGAGGCATATCCAGAGGAAGCCTACATTGCAGACCTCGATGCCAAAGTGGGGC AAGCCTGAAGCTGACCTTGCTGAACCCCAAGGGAGGATCTGGACCATGGTGGCCGGG GGTGGCGCCTCTGTCTGTACAGCGATACCATCTGTGATCTAGGGGGTGTCAACGAGC TGGCAAATATGGGGAGTACTCAGGCGCCCCCAGCGAGCAGCAGACCTATGATTATGC CAAGACTATCCTCTCCCTCATGACCCGAGAGAAGCACCAGATGGCAAGATCCTCATC ATTGGAGGCAGCATCGCAAACCTTCAACCAACGTGGCTGCCACGTTCAAGGGCATCGTGA GAGCAATTCGAGATTACCAGGGCCCCCTGAAGGAGCAGCAAGTCACAATCTTTGTCG AAGAGGTGGCCCCAACTATCAGGAGGGCTTACGGGTGATGGGAGAAGTCGGGAAGACC ACTGGGATCCCCATCCATGTCTTTGGCACAGAGACTCATATGACGGCCATTGTGGGCA TGGCCCTGGGCCACCGGCCATCCCCAACCCAGCCACCCACAGCGGCCACACTGCAAAA CTTCTCTCTCAACGCCAGCGGGAGCACATCGACGCCAGCCCCAGCAGGACAGCATCT TTTTCTGAGTCCAGGGCCGATGAGGTGGCGCTGCAAAGAAGGCCAAGCCTGCCATGC CACAAGATTAGTCCCAAGTCCAAGATCCCTGCAAGGAAAGAGCACCACCTCTTCAG CCGCCACACCAAGGCCATTGTGTGGGGCATGCAGACCCGGGCCGTGCAAGGCATGCTG GACTTTGACTATGTCTGTCTCCGAGACGAGCCCTCAGTGGCTGCCATGGTCTACCCCT TCACTGGGGACCAAGCAGAAGTTTACTGGGGGCACAAAGAGATCCTGATCCCTGT CTTCAAGAACATGGCTGATGCCATGAGGAAGCACCAGGAGGTAGATGTGCTCATCAAC TTTGCTCTCTCCGCTCTGCCTATGACAGCACCATGGAGACCATGAACATATGCCCAGA TCCGACCATCGCCATCATAGCTGAAGGCATCCCTGAGGCCCTCAGGAGAAAGCTGAT CAAGAAGGCGGACCAGAAGGGAGTGACCATCATCGGACCTGCCACTGTTGGAGGCATC AAGCCTGGGTGCTTTAAGATTGGCAACACAGGTGGGATGCTGGACAACATCCTGGCCT CCAAACGTGTACCGCCAGGCCAGCGTGGCCTATGTCTCAGTTCCGAGGCATGTCCAA CGAGCTCAACAATATCATCTCTCGGACCACGGATGGCGTCTATGAGGGCGTGGCCATT GGTGGGGACAGGTACCCGGGCTCCACATTCATGGATCATGTGTTACGCTATCAGGACA CTCCAGGAGTCAAAATGATTGTGGTCTTGGAGAGATTGGGGGCACTGAGGAATATAA GATTGCGGGGCATCAAGGAGGGCCGCTCACTAAGCCCATCGTCTGCTGGTGCATC GGGACGTGTGCCACCATGTTCTCTCTGAGGTCCAGTTTGGCCATGCTGGAGCTTGTG CCAACCAGGCTTCTGAACTGCAGTAGCCAAGAACCAGGCTTTGAAGGAAGCAGGAGT GTTTGTGCCCCGAGCTTTGATGAGCTTGGAGAGATCATCCAGTCTGTATACGAAGAT CTCGTGGCCAATGGAGTCATTGTACCTGCCAGGAGGTGCCGCCCAACCGTGCCCA TGGACTACTCTGGGCCAGGGAGCTTGGTTTGATCCGCAAACCTGCCTCGTTTCATGAC CAGCATCTGCGATGAGCGAGGACAGGAGCTCATCTACGCGGGCATGCCATCACTGAG GTCTTCAAGGAAGAGATGGGCATTGGCGGGTCTCGGCCTCTCTGGTTCCAGAAAA GGTGCTTAAGTACTCTTGCCAGTTTCAATTGAGATGTGTCTGATGGTGACAGCTGATCA CGGGCCAGCCGTCTCTGGAGCCCAACACCATCATTTGTGCGCGAGCTGGGAAAGAC CTGGTCTCCAGCCTCACCTCGGGGTGCTCACCATCGGGGATCGGTTTGGGGGTGCCT TGGATGCAGCAGCCAAGATGTTTCAAGTAAAGCCTTTGACAGTGGCATTATCCCCATGGA GTTTGTGAACAAGATGAAGAAGGAAGGAAGCTGATCATGGGCATTGGTCACCGAGTG AAGTCGATAAACAACCCAGACATGCGAGTGCGAGATCCTCAAGATTACGTCAGGCAGC ACTTCCCTGCCACTCCTCTGCTCGATTATGCACTGGAAGTAGAGAAGATTACCACCTC		

	GAAGAAGCCAAATCTTATCCTGAATGTAGATGGTCTCATCGGAGTCGCATTGTAGAC ATGCTTAGAACTGTGGGTCCCTTACTCGGGAGGAAGCTGATGAATATATTGACATTG GAGCCCTCAATGGCATCTTTGTGCTGGGAAGGAGTATGGGGTTCATTGGACACTATCT TGATCAGAAGAGGCTGAAGCAGGGGCTGTATCGTCATCCGTGGGATGATATTTTCATAT GTTCTTCCGGAACACATGAGCATGCATCATCACCACCATCACTAAGCGGCCGCTTTTCG AATC		
	ORF Start: at 1		ORF Stop: TAA at 3349
	SEQ ID NO: 250	1116 aa	MW at 122570.8kD
NOV15m, 296463359 Protein Sequence	PGPKRADSTMSAKAISEQTGKELLYKFICTTSAIQNRFKYARVTPDWDWARLLQDHPW LLSQNLVVVKPDQLIKRRGKLGVLGVNLTLDGVKSWLKPRLGQEAATVGKATGFLKNFLI EPFVPHSQAEFVYCIYATREGDYVLFHHEGGVDVGDVDAKAQKLLVGVDEKLNPEDI KKHLLVHAPEDKKEILASFISGLFNFYEDLYFTYLEINPLVVTKDGVYVLDLAAKVDA TADYICKVKWGDIEFPFPFGREAYPEEAYIADLDAKSGASLKLTLNPKGRIWTMVG GGASVVYSDTICDLGGVNELANYGEYSGAPSEQQTYDYAKTILSLMTREKHPDGKILI IGGSIANFTNVAATFKGIVRAIRDYQGPKLKEHEVTIFVRRGGPNYQEGLRVMGEVGKT TGIPIHVFGTETHMTAIVGMALGHRPIPNQPPTAAHTANFLNLSGTSSTPAPSRTAS FSESRADEVAPAKKAKPAMPQDSVPSRSLQKSTTLFSRHTKAIWGMQTRAVQGM LDFDYVCSRDEPSVAAMVYPPTGDHKQKFYWGHEKILIPVFKNMADAMRKHPEVDVLN FASLSAYDSTMETMNYAQIRTIATIAEGIPALTRKLIKADQKGVTIIGPATVGGI KPGCFKIGNTGMLDNILASKLYRPGSVAVVSRSGGMSNELNNIISRITDGVYEGVAI GGDRYPGSTFMDHVLRYQDTPGVKMIVVLGEIGGTEEYKICRGIKEGRLTPIVCWCI GTCATMFSSEVQFGHAGACANQASETAVAKNQALKEAGVFVPRSFDLGEIIQSVDYED LVANGVIVPAQEVPPPTVPMDSWARELGLIRKPASFMTSICDERGQELIYAGMPITE VFKEEMGIGGVLGLLWFQKRLPKYSCQFIEMCLMVTADHGPVSGAHTNTICARAGKD LVSSLTSGLLTIGDRFGGALDAAAKMFSKAFDSGIIPEFVNKMKKEGKLIMGIGHRV KSINNPDMRVQILKDYVRQHPATPLLDYALEVEKITTSKKPNLILNVDLIGVAFVD MLRNCGSFRTREEDEYIDIGALNGIFVLGRSMGFIGHYLDQKRLKQGLYRHPWDDISY VLPEHMSMHHHHHH		
	SEQ ID NO: 251	3313 bp	
NOV15n, 263470992 DNA Sequence	TTCCACCATGTCGGCCAAGGCAATTTCAAGACAGACGGGCAAGAAGCTCCTTTACAG TTCATCTGTACCACCTCAGCATCCAGAATCGGTTCAAGTATGCTCGGGTCACTCCTG ACACAGACTGGGCCCGCTTGTGTCAGGACCAACCCCTGGCTGCTGACGCAAGACTTGT AGTCAAGCCAGACCAGCTGATCAAACGTCGTGGAAAACCTGGTCTCGTTGGGGTCAAC CTCACTCTGGATGGGGTCAAGTCTCGGCTGAAGCCACGGCTGGGACAGGAAGCCACAG TTGGCAAGGCCACAGGCTTCCTCAAGAACCTTCTGATCGAGCCCTTCGTCCCCACAG TCAGGCTGAGGAGTCTATGTCTGCATCTATGCCACCCGAGAAGGGGACTACGTCCTG TTCCACCAAGGAGGGGTGTGACGTCGGGTGATGTGACGCCAAGCCGAGAGCTGCTG TTGTTGGCGTGGATGAGAACTGAATCCTGAGGACATCAAAAAACACCTGTTGGTCCA CGCCCTGAAGACAAGAAAGAAATTCTGGCCAGTTTTATCTCCGGCTCTTCAATTTT TACGAGGACTTGTACTTCACCTACCTCGAGATCAATCCCTTGTAGTGACCAAAGATG GAGTCTATGTCTCTGACTTGGCGGCCAAGGTGGACGCCACTGCCACTACATCTGCAA AGTGAAGTGGGGTGACATCGAGTTCCTCCCCCTTCGGGCGGGAGGCATATCCAGAG GAAGCCTACATGTCAGACCTCGATGCCAAAAGTGGGGCAAGCCTGAAGCTGACCTTGC TGAACCCCAAGGGAGGATCTGGACCATGGTGGCCGGGGTGGCGCCTCTGTCTGTGTA CAGCGATACCATCTGTGATCTAGGGGGTGTCAACGAGCTGGCAAACATATGGGGAGTAC TCAGGCGCCCCAGCGAGCAGCAGACCTATGATTATGCCAAGACTATCCTCTCCCTCA TGACCCGAGAGAAGCACCAGATGGCAAGATCCTCATATTGGAGGCAGCATCGCAAA CTTACCAACGTGGCTGACGTTCAAGGCATCGTGAGAGCAATTCAGATTACAG GGCCCCCTGAAGGAGCACGAAGTCACAATCTTTGTCCGAAGAGGTGGCCCCAACTATC AGGAGGGCTTACGGGTGATGGGAGAAGTCGGGAAGACCACTGGGATCCCCATCCATGT CTTTGGCACAGAGACTCACATGACGGCCATTGTGGGCATGGCCCTGGGCCACCGGCCC ATCCCCAACGACCCACAGCGGCCACACTGCAAACTTCTCTCAACGCCAGCG GGAGCACATCGACGCCAGCCCCAGCAGGACAGCATCTTTTCTGAGTCCAGGGCCGA TGAGGTGGCGCTGCAAGAAGGCCAAGCCTGCCATGCCACAAGATTCAAGTCCCAAGT CCAAGATCCCTGCAAGGAAAGAGCACCACCTCTTCAAGCCGACACCAAGGCCATTG TGTGGGGCATGCAGACCCGGGCCGTGCAAGGCATGTGGACTTTGACTATGTCTGCTC CCGAGACGAGCCCTCAGTGGCTGCCATGCTTACCTTTCACTGGGGACCAAGCAG AAGTTTTACTGGGGGCACAAAGAGATCCTGATCCCTGTCTTCAAGAACATGGCTGATG		

	<p>CCATGAGGAAGCACCCGAGGTAGATGTGCTCATCAACTTTGCTCTCTCCGCTCTGC  CTATGACAGCACCATGGAGACCATGAACTATGCCAGATCCGGACCATCGCCATCATA  GCTGAAGGCATCCCTGAGGCCCTCACGAGAAAGCTGATCAAGAGGCGGACCAGAAGG  GAGTGACCATCATCGGACCTGCCACTGTTGGAGGCATCAAGCCTGGGTGCTTAAAGAT  TGGCAACACAGGTGGGATGCTGGACAACATCCTGGCCTCCAACTGTACCGCCAGGC  AGCGTGGCTATGTCTCAGTTCCGGAGGCATGTCCAACGAGCTCAACAATATCATCT  CTCGGACCACGGATGGCGTCTATGAGGGCGTGGCCATTGGTGGGGACAGGTACCCGGG  CTCCACATTCATGGATCATGTGTACGCTATCAGGACACTCCAGGAGTCAAATGATT  GTGGTTCTTGGAGAGATTGGGGGCACTGAGGAATATAAGATTGCGGGGGCATCAAGG  AGGGCCGCTCACTAAGCCCATCGTCTGCTGGTGCATCGGGACGTGTGCCACCATGTT  CTCCTCTGAGGTCCAGTTTGGCCATGCTGGAGCTTGTGCCAACAGGCTTCTGAAACT  GCAGTAGCCAAGAACCAGGCTTTGAAGGAAGCAGGAGTGTGTGCCCCGGAGCTTTG  ATGAGCTTGGAGAGATCATCCAGTCTGTATACGAAGATCTCGTGGCCAATGGAGTCAT  TGTACCTGCCAGGAGGTGCCGCCCAACCGTGGCCATGGACTACTCCTGGGCCAGG  GAGCTTGGTTTGATCCGCAAACCTGCCTCGTTCATGACCAGCATCTGCGATGAGCGAG  GACAGGAGCTCATCTACGCGGGCATGCCCATCACTGAGGTCTTCAAGGAAGAGATGGG  CATTGGCGGGGTCTCGGCCTCCTCTGGTTCCAGAAAAGGTTGCCTAAGTACTCTTGC  CAGTTCATTGAGATGTGTCTGATGGTGACAGCTGATCACGGGCCAGCCGTCTCTGGAG  CCCACAACACCATCATTTGTGCGCGAGCTGGGAAAGACCTGTCTCCAGCCTCACCTC  GGGGTGCTCACCATCGGGGATCGGTTTGGGGGTGCCTTGGATGCAGCAGCCAAGATG  TTCAGTAAAGCCTTTGACAGTGGCATTATCCCCATGGAGTTTGTGAACAAGATGAAGA  AGGAAGGGAAGCTGATCATGGGCATTGGTCACCGAGTGAAGTCGATAAAACAACCCAGA  CATGCGAGTGCAGATCCTCAAAGATTACGTCAGGCAGCACTTCCCTGCCACTCCTCTG  CTCGATTATGCACCTGGAAGTAGAGAAGATTACACCTCGAAGAAGCCAAATCTTATCC  TGAATGTAGATGGTCTCATCGGAGTCGCATTTGTAGACATGCTTAGAACTGTGGGTC  CTTTACTCGGGAGGAAGCTGATGAATATATTGACATTGGAGCCCTCAATGGCATCTTT  GTGCTGGGAAGGAGTATGGGGTTCATTGGACACTATCTTGATCAGAAGAGGCTGAAGC  AGGGGCTGTATCGTCATCCGTGGGATGATATTTTCATATGTTCTTCCGGAACACATGAG  CATGTAA</p>
	<p>ORF Start: at 2</p> <p>ORF Stop: TAA at 3311</p>
	<p>SEQ ID NO: 252</p> <p>1103 aa</p> <p>MW at 121026.1kD</p>
<p>NOV15n, 263470992 Protein Sequence</p>	<p>STMSAKAI SEQTGKELLYKFICTTSAIQNRFKYARVTPD TDWARLLQDHPWLLSQNLV  VKPDQLIKRRGKLGVLGVNLTLDGVKSWLKPRLGQEATVGKATGFLKNFLIEPFVPHS  QAE EFYVCIYATREGDYVLFHHEGGVDVGDVDAKAQKLLVGVDEKLNPEDIKKHLLVH  APEDKKEILASFISGLNFYEDLYFTYLEINPLVVTKDGVVLDLAAKV DATADYICK  VKWGDIEFPFPFGREAYPEEAYIADLDAKSGASLKLTLNPKGRIWTMVAGGGASVVY  SDTICDLGGVNELANYGEYS GAPSEQTYDYAKTILSLMTREKHPDGKILIGGSIAN  FTNVAATFKGI VRAIRDYQG PLKEHEVTI FVRGGPNYQELGRVMEVGKTTGIP I HV  FGTETHMTAIVGMALGHRPIPNQPPTAHTANFLLNASGSTSTPAPSR TASFSERAD  EVAPAKKAKPAMPQDSVSPRSLQKSTTLFSRHTKAI VWGMQTRAVQGM LDFDYVCS  RDEPSVAAMVYPFTGDHKQKFYWGHEKILIPVFKNMADAMRKHP EVDVLINFASLRS  YDSTMETMNYAQIRTI AIAEGIPEALTRKLIKADQKGV TIIGPATVGGIKPGCFKI  GNTGGMLDNILASKLYRPGSVAYVSRSGMSNELNNIISRTDGVYEGVAIGGDRYPG  STFMDHVLRYQDTPGVKMIVVLGEIGGTEEYKICRG IKEGRLTKPIVCWIGTCATMF  SSEVQFGHAGACANQASETAVAKNQALKEAGVFVPRSFDELGEIIQS VYEDLVANGVI  VPAQEVPPPTVPM DYSWARELGLIRKPASFM TSICDERGQELIYAGMPI TEVFKEEMG  IGGVLG LLWFQKRLPKYSCQFIEMCLMVTADHGP AVSGAHTIIICARAGKDLVSSLT  SGLLTIGDRFGGALDAAAKMFSKAFDSGII PMEFVNKMKKEGKLIMG IHRVKSINNP  MRVQILKDYVRQHF PATPLLDYALEVEKITTSKPNLILNVDLIGV FVDM L RNCGS  FTREEADEYIDIGALNGIFVLGRSMGFIGHYLDQKRLKQGLYRHPWDDISYVLPEHMS  M</p>
	<p>SEQ ID NO: 253</p> <p>3368 bp</p>
<p>NOV15o, CG142427-05 DNA Sequence</p>	<p>CCCGTCCGAAGCGCGCGGATTCCACCATGTCGGCCAAGGCAATTCAGAGCAGACGG  GCAAAGAACTCCTTTACAAGTTCATCTGTACCACCTCAGCCATCCAGAATCGGTTCAA  GTATGCTCGGGTCACTCCTGACACAGACTGGGCCCGCTTGCTGCAGGACCACCCCTGG  CTGCTCAGCCAGAACTTGGTAGTCAAGCCAGACCAGCTGATCAAACGTCGTGGAAAC  TTGGTCTCGTTGGGGTCAACCTCACTCTGGATGGGGTCAAGTCCTGGCTGAAGCCACG  GCTGGGACAGGAAGCCACAGTTGGCAAGGCCACAGGCTTCTCAAGAACTTTCTGATC</p>

	GAGCCCTTCGTCCCCACAGTCAGGCTGAGGAGTTCTATGTCTGCATCTATGCCACCC GAGAAGGGGACTACGTCTCTTCCACCACGAGGGGGTGTGGACGTGGGTGATGTGGA CGCCAAGGCCCAGAAGCTGCTTGTGGCGTGGATGAGAACTGAATCCTGAGGACATC AAAAAACACCTGTTGGTCCACGCCCCGAAGACAAGAAAGAAATCTGGCCAGTTTTA TCTCCGGCCTCTTCAATTTCTACGAGGACTTGTACTTCACCTACCTCGAGATCAATCC CCTTGTAGTGACCAAAGATGGAGTCTATGTCTTGACTTGGCGGCCAAGGTGGACGCC ACTGCCGACTACATCTGCAAAGTGAAGTGGGGTGACATCGAGTTCCCTCCCCCTTCG GGCGGGAGGCATATCCAGAGGAAGCCTACATTGCAGACCTCGATGCCAAAAGTGGGGC AAGCCTGAAGCTGACCTTGCTGAACCCCAAAGGGAGGATCTGGACCATGGTGGCCGGG GGTGGCGCCTCTGTCTGTACAGCGATACCATCTGTGATCTAGGGGGTGTCAACGAGC TGGCAAACCTATGGGGAGTACTCAGGCGCCCCCAGCGAGCAGACAGCATATGATTATGC CAAGACTATCCTCTCCCTCATGACCCGAGAGAAGCACCAGATGGCAAGATCCTCATC ATTGGAGGCAGCATCGCAAACCTTACCAACGTGGCTGCCACGTTCAAGGGCATCGTGA GAGCAATTGAGATTACAGGGCCCCCTGAAGGAGCAGCAAGTCACAATCTTTGTCCG AAGAGGTGGCCCCAACTATCAGGAGGGCTTACGGGTGATGGGAGAAGTCGGGAAGACC ACTGGGATCCCCATCATGTCTTTGGCACAGAGACTCACATGACGGCCATTGTGGGGCA TGGCCCTGGGCCACCGGCCATCCCCAACGAGCCACCCACAGCGGCCACACTGCAAA CTTCTCTCAACGCCAGCGGGAGCACATCGACGCCAGCCCCCAGCAGGACAGCATCT TTTTCTGAGTCCAGGGCCGATGAGGTGGCGCTGCAAAGAAGGCCAAGCCTGCCATGC CACAAGATTAGTCCCAAGTCCAAGATCCCTGCAAGGAAAGAGCACCACCTCTTCAG CCGCCACACCAAGGCCATTGTGTGGGGCATGCAGACCCGGGCCGTGCAAGGCATGCTG GACTTTGACTATGTCTGCTCCCGAGACGAGCCCTCAGTGGCTGCCATGGTCTACCCCT TCACTGGGGACCACAAGCAGAAGTTTTACTGGGGGCACAAAGAGATCCTGATCCCTGT CTTCAAGAACATGGCTGATGCCATGAGGAAGCACC CGGAGGTAGATGTGCTCATCAAC TTTGCCCTCTCTCCGCTCTGCCTATGACAGCACCATGGAGACCATGAACATATGCCCAGA TCCGGACCATCGCCATCATAGCTGAAGGCATCCCTGAGGCCCTCACGAGAAAGCTGAT CAAGAAGGCGGACCAGAGGGAGTGACCATCATCGGACCTGCCACTGTTGGAGGCATC AAGCCTGGGTGCTTTAAGATTGGCAACACAGGTGGGATGCTGGACAACATCCTGGCCT CCAAACTGTACCGCCCAGGCAGCGTGGCCTATGTCTCACGTTCCGGAGGCATGTCCAA CGAGCTCAACAATATCATCTCTCGGACCACGGATGGCGTCTATGAGGGCGTGGCCATT GGTGGGGACAGGTACCCGGGCTCCACATTTCATGGATCATGTGTTACGCTATCAGGACA CTCCAGGAGTCAAAATGATTGTGGTTCTTGGAGAGATTGGGGGCACTGAGGAATATAA GATTTGCCGGGGCATCAAGGAGGGCCGCTCACTAAGCCCATCGTCTGCTGGTGATC GGGACGTGTGCCACCATGTTCTCTCTGAGGTCCAGTTTGGCCATGCTGGAGCTTGTG CCAACCAGGCTTCTGAAACTGCAGTAGCCAAGAACCAGGCTTTGAAGGAAGCAGGAGT GTTTGTGCCCCGAGCTTTGATGAGCTTGGAGAGATCATCCAGTCTGTATACGAAGAT CTCGTGGCCAATGGAGTCATTGTACCTGCCCAGGAGGTGCCGCCCCAACCGTGCCCA TGGACTACTCCTGGGCCAGGGAGCTTGGTTTGATCCGCAACCTGCCTCGTTTCATGAC CAGCATCTGCGATGAGCGAGGACAGGAGCTCATCTACGCGGGCATGCCCATCACTGAG GTCTTCAAGGAAGAGATGGGCATTGGCGGGGTCTCGGCCCTCCTCTGGTTCCAGAAAA GGTTGCCTAAGTACTCTTGCCAGTTCATTGAGATGTGTCTGATGGTGACAGCTGATCA CGGGCCAGCGTCTCTGGAGCCCAACACCATCATTTGTGCGCGAGCTGGGAAAGAC CTGGTCTCCAGCCTCACCTCGGGGCTGCTCACCATCGGGGATCGGTTTGGGGGTGCCT TGGATGCAGCAGCCAAGATGTTTCAGTAAAGCCTTTGACAGTGGCATTATCCCCATGGA GTTTGTGAACAAGATGAAGAAGGAAGGGAAGCTGATCATGGGCATTGGTCAACCGAGTG AAGTCGATAAACAACCCAGACATGCGAGTGCAGATCCTCAAAGATTACGTCAGGCAGC ACTTCCCTGCCACTCCTCTGCTCGATTATGCACTGGAAGTAGAGAAGATTACCACCTC GAAGAAGCCAAATCTTATCCTGAATGTAGATGGTCTCATCGGAGTCGCATTTGTAGAC ATGCTTAGAAACTGTGGGTCTTTACTCGGGAGGAAGCTGATGAATATATTGACATTG GAGCCCTCAATGGCATCTTTGTGCTGGGAAGGAGTATGGGGTTTATTGGACACTATCT TGATCAGAAGAGGCTGAAGCAGGGGCTGTATCGTCATCCGTGGGATGATATTTTCATAT GTTCTTCCGGAACACATGAGCATGCATCATCACCACCATCACTAAGCGGCCGCTTTCG AATC		
	ORF Start: ATG at 28		ORF Stop: at 3331
	SEQ ID NO: 254	1101 aa	MW at 120838.0kD
NOV15o, CG142427-05 Protein Sequence	MSAKAISEQTGKELLYKFICTTSAIQNRFKYARVTPDIDWARLLQDHPWLLSQNLVVK PDQLIKRRGKGLGVGNLTLDGVKSWLKPRLGQEATVGKATGFLKNFLIEFPVPHSQA EEFYVCIYATREGDYVLFHHEGGVDVGDVDAKAQKLLVGVDEKLPEDIKKHLVHAP EDKKEILASFISGLNFYEDLYFTYLEINPLVVTKDGVYVLDLAAKV DATADYICKVK		

	WGDIEFPPPPGREAYPEEAYIADLDAKSGASLKLTLNPKGRIWTMVAGGGASVVYSD TICDLGGVNELANYGEYS GAPSEQQTYDYAKTILSLMTRKHPDGKILIIIGGSIANFT NVAATFKGIVRAIRDYQGPLKEHEVTIFVRRGGPNYQEGLRVMGEVGKTTGIPIHVFG TETHMTAIVGMALGHRPIPNQPPTAAHTANFLLNASGSTSTPAPSRITASFSESRADDEV APAKKAKPAMPQDSVSPSRSLQKSTTLFSRHTKAIWGMQTRAVQGMDFDYVCSR EPSVAAMVYPFTGDHKQKFWGHKEILIPVFKNMADAMRKHPVDVLINFASLSAYD STMETMNYAQIRTIATIAEGIPALTRKLIKADQKGVTTIIGPATVGGIKPGCFKIGN TGGMLDNILASKLYRPGSVAYVSRSGGMSNELNNIISRTTDGVYEGVAIGGDRYPGST FMDHVLRYQDTPGVKMIIVVLGEIGTEEYKICRGIKEGRLTKPIVCWICGTCATMFSS EVQFGHAGACANQASETAVAKNQALKEAGVFVPRSPFDELGEIIQSVYEDLVANGVIVP AQEVPPPTVPMDYWARELGLIRKPASFMSTICDERGQELIYAGMPITEVFKEEMGIG GVLGLLWFQKRLPKYSCQFIEMCLMVTADHGPAVSGAHNTIICARAGKDLVSSLTSG LTIGDRFGGALDAAAKMFSKAFDSGIIPMEFVNKMKKEGKLIMGIGHRVKSINNPDMR VQILKDYVRQHFPAATPLLDYALEVEKITTSKKPNLILNVDGLIGVAFVDMRLNCGSFT REEAD EYIDIGALNGIFVLGRSMGFIGHYLDQKRLKQGLYRHPWDDISYVLP EHMMSM
--	--

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 15B.

<b>Table 15B. Comparison of NOV15a against NOV15b through NOV15o.</b>		
<b>Protein Sequence</b>	<b>NOV15a Residues/ Match Residues</b>	<b>Identities/ Similarities for the Matched Region</b>
NOV15b	1..1101 1..1101	1101/1101 (100%) 1101/1101 (100%)
NOV15c	1..1101 5..1072	1065/1101 (96%) 1065/1101 (96%)
NOV15d	1..1101 5..1095	1091/1101 (99%) 1091/1101 (99%)
NOV15e	1..589 5..604	570/610 (93%) 573/610 (93%)
NOV15f	1..1101 5..1105	1101/1101 (100%) 1101/1101 (100%)
NOV15g	1..1101 5..1095	1091/1101 (99%) 1091/1101 (99%)
NOV15h	1..589 5..604	570/610 (93%) 573/610 (93%)
NOV15i	1..1101 2..1102	1101/1101 (100%) 1101/1101 (100%)
NOV15j	1..1101 5..1105	1101/1101 (100%) 1101/1101 (100%)
NOV15k	1..1101 5..1095	1091/1101 (99%) 1091/1101 (99%)
NOV15l	1..589 5..604	570/610 (93%) 573/610 (93%)
NOV15m	1..1101	1101/1101 (100%)

	10..1110	1101/1101 (100%)
NOV15n	1..1101	1101/1101 (100%)
	3..1103	1101/1101 (100%)
NOV15o	1..1101	1101/1101 (100%)
	1..1101	1101/1101 (100%)

Further analysis of the NOV15a protein yielded the following properties shown in Table 15C.

Table 15C. Protein Sequence Properties NOV15a	
PSort analysis:	0.8500 probability located in endoplasmic reticulum (membrane); 0.4450 probability located in microbody (peroxisome); 0.4400 probability located in plasma membrane; 0.1000 probability located in mitochondrial inner membrane
SignalP analysis:	No Known Signal Sequence Predicted

5. A search of the NOV15a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 15D.

Table 15D. Geneseq Results for NOV15a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV15a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
ABB61832	Drosophila melanogaster polypeptide SEQ ID NO. 12288 - Drosophila melanogaster, 1086 aa. [WO200171042-A2, 27-SEP-2001]	1..1097 1..1083	762/1099 (69%) 895/1099 (81%)	0.0
AAB56952	Human prostate cancer antigen protein sequence SEQ ID NO:1530 - Homo sapiens, 363 aa. [WO200055174-A1, 21-SEP-2000]	753..1101 15..363	347/349 (99%) 347/349 (99%)	0.0
AAY67408	Arabidopsis ATP citrate lyase (ACL) B-2 subunit - Arabidopsis sp, 608 aa. [WO200000619-A2, 06-JAN-2000]	492..1093 6..606	321/602 (53%) 429/602 (70%)	0.0
AAG36247	Arabidopsis thaliana protein fragment SEQ ID NO: 44394 - Arabidopsis thaliana, 681 aa. [EP1033405-A2, 06-SEP-2000]	492..1093 6..606	321/602 (53%) 429/602 (70%)	0.0

AAG36248	Arabidopsis thaliana protein fragment SEQ ID NO: 44395 - Arabidopsis thaliana, 656 aa. [EP1033405-A2, 06-SEP-2000]	512..1093 1..581	313/582 (53%) 417/582 (70%)	0.0
----------	--	---------------------	--------------------------------	-----

In a BLAST search of public sequence databases, the NOV15a protein was found to have homology to the proteins shown in the BLASTP data in Table 15E.

Table 15E. Public BLASTP Results for NOV15a				
Protein Accession Number	Protein/Organism/Length	NOV15a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
P53396	ATP-citrate (pro-S-)-lyase (EC 4.1.3.8) (Citrate cleavage enzyme) - Homo sapiens (Human), 1101 aa.	1..1101 1..1101	1100/1101 (99%) 1101/1101 (99%)	0.0
P16638	ATP-citrate (pro-S-)-lyase (EC 4.1.3.8) (Citrate cleavage enzyme) - Rattus norvegicus (Rat), 1100 aa.	1..1101 1..1100	1074/1101 (97%) 1086/1101 (98%)	0.0
Q91V92	ATP-citrate (pro-S-)-lyase (EC 4.1.3.8) (Citrate cleavage enzyme) - Mus musculus (Mouse), 1091 aa.	1..1101 1..1091	1070/1101 (97%) 1083/1101 (98%)	0.0
S21173	ATP citrate (pro-S)-lyase - human, 1105 aa.	1..1101 1..1105	1078/1106 (97%) 1082/1106 (97%)	0.0
Q8VIQ1	ATP-citrate lyase - Rattus norvegicus (Rat), 851 aa (fragment).	250..1101 1..851	835/852 (98%) 842/852 (98%)	0.0

PFam analysis predicts that the NOV15a protein contains the domains shown in the Table 15F.

Table 15F. Domain Analysis of NOV15a			
Pfam Domain	NOV15a Match Region	Identities/ Similarities for the Matched Region	Expect Value
CoA_binding	492..616	33/126 (26%) 88/126 (70%)	1.5e-19
ligase-CoA	642..793	49/156 (31%) 126/156 (81%)	3.9e-53

## 5 Example 16.

The NOV16 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 16A.

Table 16A. NOV16 Sequence Analysis			
	SEQ ID NO: 255	1393 bp	
NOV16a, CG142631-01 DNA Sequence	CCTTCTCTTCGTGGGCTATCTACTCAGTTGATCCCTCCCTCGCTGGCTTGGCTCTGAC TCCTGCTCAGACCCATCACCTTTGCCGGGAATGATGTCTGGAGAACCCCTGCACGTG AAGACCCCATCCGTGACAGCATGGCCCTGTCCAAATGGCCGGCACCAGCGTCTACC TCAAGATGGACAGTGGCCAGCCCTCCGGCTCCTTCAAGATCCGGGGCATTGGGCACTT CTGCAAGAGGTGGGCCAAGCAAGGCTGTGCACATTTTGTCTGCTCCTCGGCGGGCAAC GCAGGCATGGCGGCTGCATATGCGGCCAGGCAACTCGGCGTCCCCGCCACCATCGTAG TGCCCGGCACCACACCTGCTCTCACCATTGAGCGCCTCAAGAATGAAGGTGCCACATG CAAGGTGGTGGGTGAGTTATTGGATGAAGCCTTCGAGCTGGCCAAGGCCCTAGCGAAG AACAACCCGGGTTGGGTCTACATTCCCCCTTTGATGACCCCTCATCTGGGAAGGCC ACGCTTCCATCGTGAAAGAGCTGAAGGAGACACTGTGGGAAAAGCCGGGGGCCATCGC GCTGTCAAGTGGGCGGCGGGGCTGCTGTGTGGAGTGGTCCAGGGGCTGCAGGAGTGT GGCTGGGGGGACGTGCCTGTCTATCGCCATGGAGACTTTTGGTGCCACAGCTTCCACG CTGCCACCACCGCAGGCAAACTTGTCTCCCTGCCAAGATCACCAGTGTGGCCAAGGC CCTGGGCGTGAAGACTGTGGGTCTCAGGCCCTGAAGCTGTTTCAGGAACACCCCAT TTCTCTGAAGTTATCTCGGACCAGGAGGCTGTGGCCGCCATTGAGAAGTTCGTGGATG ATGAGAAGATCCTGGTGGAGCCCGCTGGGCGCAGCCCTGGCCGCTGTCTATAGCCA CGTGATCCAGAAGCTCCAATGGAGGGGAATCTCCGAACCCCGCTGCCATCCCTCGTG GTCATCGTCTGCGGGGGCAGCAACATCAGCCTGGCCAGCTGCGGGCGCTCAAGGAAC AGCTGGGCATGACAAATAGGTTGCCCAAGTGAGGACGACCCCTTACCGATCTGTGCT CTCCTAGCCCAAGAGACCCCTGGAGGGGCTGGAGTTTATCCAGCGCTCGTCTATGT TTGGCTGAGCACCTGTGGCCCTGGGTGCAGGTTAACTTCTTGTATCAGGAGCCCACT ATGCAGAGGCCAAAGGTTCGGCAGCCAGCGAGGCTATGAATTGGACCTTTTGGTATCT GTGTGACTGCTCTGTGCCCATCCTTAGCCAACTTGCTGGCGTGACAAGTGCCCACAAG TAACACACCAGGTACCCAGAGCAGGGTGGACAGGAGAGACCTGAATCACAGCAGTGAG G		
	ORF Start: ATG at 90		ORF Stop: TGA at 1074
	SEQ ID NO: 256	328 aa	MW at 34702.1kD
NOV16a, CG142631-01 Protein Sequence	MMSGEPLHVKTPIRDSMALSKMAGTSVYLKMDSAQPSGSFKIRGIGHFKRWAKQGCA HFVCSAGNAGMAAAYAARQLGVPATIVPGTTPALTIERLKNEGATCKVVGELLDEA FELAKALAKNNPGWYIIPFDDPLIWEHGASIVKELKETLWEKPGAIALSVGGGGLLC GVVQGLQECGWGDVPVIAMETFGAHSFHAATTAGKLVS LPKITSVAKALGVKTVGSQA LKL FQEHPIFSEVISDQEA VAAIEKFVDEKILVEPAWGAALAAVYSHVIQKLQLEGN LRTPLPSLVVIVCGGSNISLAQLRALKEQLGMTNRLPK		
	SEQ ID NO: 257	1393 bp	
NOV16b, CG142631-01 DNA Sequence	CCTTCTCTTCGTGGGCTATCTACTCAGTTGATCCCTCCCTCGCTGGCTTGGCTCTGAC TCCTGCTCAGACCCATCACCTTTGCCGGGAATGATGTCTGGAGAACCCCTGCACGTG AAGACCCCATCCGTGACAGCATGGCCCTGTCCAAATGGCCGGCACCAGCGTCTACC TCAAGATGGACAGTGGCCAGCCCTCCGGCTCCTTCAAGATCCGGGGCATTGGGCACTT CTGCAAGAGGTGGGCCAAGCAAGGCTGTGCACATTTTGTCTGCTCCTCGGCGGGCAAC GCAGGCATGGCGGCTGCATATGCGGCCAGGCAACTCGGCGTCCCCGCCACCATCGTAG TGCCCGGCACCACACCTGCTCTCACCATTGAGCGCCTCAAGAATGAAGGTGCCACATG CAAGGTGGTGGGTGAGTTATTGGATGAAGCCTTCGAGCTGGCCAAGGCCCTAGCGAAG AACAACCCGGGTTGGGTCTACATTCCCCCTTTGATGACCCCTCATCTGGGAAGGCC ACGCTTCCATCGTGAAAGAGCTGAAGGAGACACTGTGGGAAAAGCCGGGGGCCATCGC GCTGTCAAGTGGGCGGCGGGGCTGCTGTGTGGAGTGGTCCAGGGGCTGCAGGAGTGT GGCTGGGGGGACGTGCCTGTCTATCGCCATGGAGACTTTTGGTGCCACAGCTTCCACG CTGCCACCACCGCAGGCAAACTTGTCTCCCTGCCAAGATCACCAGTGTGGCCAAGGC CCTGGGCGTGAAGACTGTGGGTCTCAGGCCCTGAAGCTGTTTCAGGAACACCCCAT TTCTCTGAAGTTATCTCGGACCAGGAGGCTGTGGCCGCCATTGAGAAGTTCGTGGATG ATGAGAAGATCCTGGTGGAGCCCGCTGGGCGCAGCCCTGGCCGCTGTCTATAGCCA CGTGATCCAGAAGCTCCAATGGAGGGGAATCTCCGAACCCCGCTGCCATCCCTCGTG GTCATCGTCTGCGGGGGCAGCAACATCAGCCTGGCCAGCTGCGGGCGCTCAAGGAAC		

	AGCTGGGCATGACAAATAGGTTTGCCCAAGTGAGGACGGACCCCTTACCGATCTGTGCT CTCCTAGCCCAAGAGACCCCTGGAGGGGCTGGAGTTTATCCAGCGCCTCGTCTGATGT TTGGCTGAGCACCTGTGGCCCTGGGTGCAGGTTAACTTCTTGTATTATCAGGAGCCCACT ATGCAGAGGCCAAAGGTCGGCAGCCAGCGAGGCTATGAATTGGACCTTTTTTGGTATCT GTGTGACTGCTCTGTGCCCATCCTTAGCCAACCTTGCTGGCGTGACAAGTGCCCAACAAG TAACACACCAGGTACCCAGAGCAGGGTGGACAGGAGAGACCTGAATCACAGCAGTGAG G		
	ORF Start: ATG at 90		ORF Stop: TGA at 1074
	SEQ ID NO: 258	328 aa	MW at 34702.1kD
NOV16b, CG142631-01 Protein Sequence	MMSGEPHVKTPIRDMSALSKMAGTSVYLKMDSAQPSGSFKIRGIGHFCKRWAKQGCA HFVCSSAGNAGMAAAYAARQLGVPATIVVPGTTPALTIERLKNEGATCKVVGELLDEA FELAKALAKNNPGWVYIPFDDPLIWEGHASIVKELKETLWEKPGAIALSVGGGGLLC GVVQGLQECGWGDPVIAMETFGAHSFHAATTAGKLVSLPKITSVAKALGVKTVGSQA LKLQFQEHPIFSEVISDQEAVAAIEKPFVDDEKILVEPAWGAALAAVYSHVIQKLQLEGN LRTPLPSLVVIVCGGSNISLAQLRALKEQLGMTNRLPK		
	SEQ ID NO: 259	1008 bp	
NOV16c, 248494617 DNA Sequence	ACCATGATGTCTGGAGAACCCCTGCACGTGAAGACCCCATCCGTGACAGCATGGCCC TGTCCAAAAATGGCCGGCACCAGCGTCTACCTCAAGATGGACAGTGCCAGCCCTCCGG CTCCTTCAAGATCCGGGGCATTGGGCACTTCTGCAAGAGGTGGGCCAAGCAAGGCTGT GCACATTTTGTCTGCTCCTCGCGGGCAACGCAGGCATGGCGGCTGCATATGCGGCCA GGCAACTCGGCGTCCCCGCCACCATCGTGGTGCCAGCACACACTGCTCTCACCAT TGAGCGCCTCAAGAATGAAGGTGCCACAGTCAAGGTGGTGGGTGAGTTATTGGATGAA GCCTTCGAGCTGGCCAAGGCCCTAGCGAAGAACAACCCGGGTTGGGTCTACATTCCCC CCTTTGATGACCCCTCATCTGGGAAGGCCACGCTTCCATCGTGAAAGAGCTGAAGGA GACACTGTGGGAAAAGCCGGGGCCATCGCGCTGTCTAGTGGGCGGCGGGGGCCTGCTG TGTGGAGTGGTCCAGGGGCTGCAGGAGGTGGGCTGGGGGGACGTGCCTGTCTATCGCCA TGGAGACTTTTGGTGCCACAGCTTCCACGCTGCCACCACCGCAGGCAAACTTGTCTC CCTGCCCAAGATCACCAGTGTGCAAGGCCCTGGGCGTGAAGACTGTGGGGGCTCAG GCCCTGAAGCTGTTTCAGGAACACCCCATTTTCTCTGAAGTTATCTCGGACCAGGAGG CTGTGGCCGCCATTGAGAAGTTCGTGGATGATGAGAAGATCCTGGTGGAGCCCGCCTG CGGGGCAGCCCTGGCCGCTGTCTATAGCCACGTGATCCAGAAGCTCCAAGTGGAGGGG AATCTCCGAACCCCGCTGCCATCCCTCGTGGTTCATCGTCTCGGGGGGCAGCAACATCA GCCTGGCCAGCTGCGGGCGCTCAAGGAACAGCTGGGCATGACAAATAGGTTGCCCAA GCATCATCACCACCATCACTGA		
	ORF Start: at 1		ORF Stop: TGA at 1006
	SEQ ID NO: 260	335 aa	MW at 35549.0kD
NOV16c, 248494617 Protein Sequence	TMSGEPHVKTPIRDMSALSKMAGTSVYLKMDSAQPSGSFKIRGIGHFCKRWAKQGC AHFVCSSAGNAGMAAAYAARQLGVPATIVVPGTTPALTIERLKNEGATVKVVGELLDE AFELAKALAKNNPGWVYIPFDDPLIWEGHASIVKELKETLWEKPGAIALSVGGGGLL CGVVQGLQEVGWGDPVIAMETFGAHSFHAATTAGKLVSLPKITSVAKALGVKTVGAQ ALKLQFQEHPIFSEVISDQEAVAAIEKPFVDDEKILVEPACGAALAAVYSHVIQKLQLEG NLRTPLPSLVVIVCGGSNISLAQLRALKEQLGMTNRLPKHHHHHH		
	SEQ ID NO: 261	988 bp	
NOV16d, 228832711 DNA Sequence	CATGATGTCTGGAGAACCCCTGCACGTGAAGACCCCATCCGTGACAGCATGGCCCTG TCCAAAAATGGCCGGCACCAGCGTCTACCTCAAGATGGACAGTGCCAGCCCTCCGGCT CCTTCAAGATCCGGGGCATTGGGCACTTCTGCAAGAGGTGGGCCAAGCAAGGCTGTGC ACATTTTGTCTGCTCCTCGCGGGCAACGCAGGCATGGCGGCTGCATATGCGGCCAGG CAACTCGGCGTCCCCGCCACCATCGTGGTGCCAGCACACACTGCTCTCACCATTG AGCGCCTCAAGAATGAAGGTGCCACAGTCAAGGTGGTGGGTGAGTTATTGGATGAAGC CTTCGAGCTGGCCAAGGCCCTAGCGAAGAACAACCCGGGTGGGTCTACATTCCCCC TTTGATGACCCCTCATCTGGGAAGGCCACGCTTCCATCGTGAAAGAGCTGAAGGAGA CACTGTGGGAAAAGCCGGGGGCCATCGCGCTGTCTAGTGGGCGGCGGGGGCCTGCTGTG TGGAGTGGTCCAGGGGCTGCAGGAGGTGGGCTGGGGGGACGTGCCTGTCTATCGCCATG GAGACTTTTGGTGCCCAAGCTTCCACGCTGCCACCACCGCAGGCAAACTTGTCTCCC TGCCAAGATCACCAGTGTGCAAGGCCCTGGGCGTGAAGACTGTGGGGGCTCAGGC CCTGAAGCTGTTTCAGGAACACCCCATTTTCTCTGAAGTTATCTCGGACCAGGAGGCT		

	GTGGCCGCCATTGAGAAGTTCGTGGATGATGAGAAGATCCTGGTGGAGCCCGCCTGCGGGCAGCCCTGGCCGCTGTCTATAGCCACGTGATCCAGAAGCTCCAACCTGGAGGGAAATCTCCGAACCCCGCTGCCATCCCTCGTGGTCATCGTCTGCGGGGGCAGCAACATCAGCCTGGCCCAGCTGCGGGCGCTCAAGGAACAGCTGGGCATGACAAATAGGTTGCCCAAGTGA		
	ORF Start: ATG at 2		ORF Stop: TGA at 986
	SEQ ID NO: 262	328 aa	MW at 34625.0kD
NOV16d, 228832711 Protein Sequence	MMSGEPLHVKTPIRDSMALSKMAGTSVYLKMDSAQPSGSFKIRGIGHFCKRWAKQGCAHFVCSSAGNAGMAAAYAARQLGVPATIVVPSTTPALTIERLKNEGATVKVVGELLDEFELAKALAKNNPGWVYIPFDDPLIWEGHASIVKELKETLWEKPGAIALSVGGGGLLCGVVQGLQEVGWGDVPVIAMETFGAHSFHAATTAGKLVS LPKITSVAKALGVKTVGAQALLKLFQEHPIFSEVISDQEA VAAIEKFVDDEKILVEPACGAALAAVYSHVIQKLQLEGNLRTPLPSLVIVCGGSNISLAQLRALKEQLGMTNRLPK		
	SEQ ID NO: 263	1035 bp	
NOV16e, 256420310 DNA Sequence	ATGTCTGGAGAACCCTGCACGTGAAGACCCCATCCGTGACAGCATGGCCCTGTCCA AAATGGCCGGCACCAGCGTCTACCTCAAGATGGACAGTGGCCAGCCCTCCGGCTCCTTCAAGATCCGGGGCATTTGGGCATTCTGCAAGAGGTGGGCCAAGCAAGGCTGTGCACATTTTGTCTGCTCCTCGGCGGGCAACGCAGGCATGGCGGCTGCATATGCGGCCAGGCAACTCGGCGTCCCCGCCACCATCGTGGTGCCAGCACCACA CTTGCTCTACCATTGAGCGCCTCAAGAATGAAGGTGCCACAGTCAAGGTGGTGGGTGAGTTATTGGATGAAGCCTTCGAGCTGGCCAAGGCCCTAGCGAAGAACAACCCGGGTTGGGTCTACATTCCCCCTTTGATGACCCCTCATCTGGGAAGGCCACGCTTCCATCGTGAAAGAGCTGAAGGAGACACTGTGGGAAAAGCCGGGGCCATCGCGCTGTCACTGGGCGGCGGGGGCCTGCTGTGTGGA GTGGTCCAGGGGCTGCAGGAGGTGGGCTGGGGGGACGTGCCTGTCTATCGCCATGGAGACTTTTGGTGCCACAGCTTCCACGCTGCCACCACCGCAGGCAAACTTGTCTCCCTGCCCAAGATCACCAGTGTGTGCAAGGCCCTGGGCGTGAAGACTGTGGGGGCTCAGGCCCTGAAGCTGTTTCAGGAACACCCCATTTTCTCTGAAGTTATCTCGGACCAGGAGGCTGTGGCCGCCATTGAGAAGTTCGTGGATGATGAGAAGATCCTGGTGGAGCCCGCTGCGGGGAGCCCTGGCCGCTGTCTATAGCCACGTGATCCAGAAGCTCCAACCTGGAGGGGAATCTCGAACCCCGCTGCCATCCCTCGTGGTCATCGTCTGCGGGGGCAGCAACATCAGCCTGGCCCAGCTGCGGGCGCTCAAGGAACAGCTGGGCATGACAAATAGGTTGCCCAAGCATCATCACCACCATCACTGAGCGGCCGCACTCGAGCACCACCACCACCACCAC		
	ORF Start: ATG at 1		ORF Stop: TGA at 1000
	SEQ ID NO: 264	333 aa	MW at 35316.7kD
NOV16e, 256420310 Protein Sequence	MSGEPLHVKTPIRDSMALSKMAGTSVYLKMDSAQPSGSFKIRGIGHFCKRWAKQGCALFVCSSAGNAGMAAAYAARQLGVPATIVVPSTTPALTIERLKNEGATVKVVGELLDEAFELAKALAKNNPGWVYIPFDDPLIWEGHASIVKELKETLWEKPGAIALSVGGGGLLCGVVQGLQEVGWGDVPVIAMETFGAHSFHAATTAGKLVS LPKITSVAKALGVKTVGAQALLKLFQEHPIFSEVISDQEA VAAIEKFVDDEKILVEPACGAALAAVYSHVIQKLQLEGNLRTPLPSLVIVCGGSNISLAQLRALKEQLGMTNRLPKHHHHH		
	SEQ ID NO: 265	1017 bp	
NOV16f, 249117058 DNA Sequence	ATGTCTGGAGAACCCTGCACGTGAAGACCCCATCCGTGACAGCATGGCCCTGTCCA AAATGGCCGGCACCAGCGTCTACCTCAAGATGGACAGTGGCCAGCCCTCCGGCTCCTTCAAGATCCGGGGCATTTGGGCATTCTGCAAGAGGTGGGCCAAGCAAGGCTGTGCACATTTTGTCTGCTCCTCGGCGGGCAACGCAGGCATGGCGGCTGCATATGCGGCCAGGCAACTCGGCGTCCCCGCCACCATCGTGGTGCCAGCACCACCTGCTCTCACCATTGAGCGCCTCAAGAATGAAGGTGCCACAGTCAAGGTGGTGGGTGAGTTATTGGATGAAGCCTTCGAGCTGGCCAAGGCCCTAGCGAAGAACAACCCGGGTTGGGTCTACATTCCCCCTTTGATGACCCCTCATCTGGGAAGGCCACGCTTCCATCGTGAAAGAGCTGAAGGAGACACTGTGGGAAAAGCCGGGGCCATCGCGCTGTCACTGGGCGGCGGGGGCCTGCTGTGTGGA GTGGTCCAGGGGCTGCAGGAGGTGGGCTGGGGGGACGTGCCTGTCTATCGCCATGGAGACTTTTGGTGCCACAGCTTCCACGCTGCCACCACCGCAGGCAAACTTGTCTCCCTGCCCAAGATCACCAGTGTGTGCAAGGCCCTGGGCGTGAAGACTGTGGGGGCTCAGGCCCTGAAGCTGTTTCAGGAACACCCCATTTTCTCTGAAGTTATCTCGGACCAGGAGGCTGTGGCCGCCATTGAGAAGTTCGTGGATGATGAGAAGATCCTGGTGGAGCCCGCTGCGGGGAGCCCTGGCCGCTGTCTATAGCCACGTGATCCAGAAGCTCCAACCTGGAGGGGAATCTC		

	CGAACCCCGCTGCCATCCCTCGTGGTCATCGTCTGCGGGGGCAGCAACATCAGCCTGG CCCAGCTGCGGGCGCTCAAGGAACAGCTGGGCATGACAAATAGGTTGCCCAAGTGAGC GGCCGCACTCGAGCACCACCACCACCACCAC		
	ORF Start: ATG at 1		ORF Stop: TGA at 982
	SEQ ID NO: 266	327 aa	MW at 34493.8kD
NOV16f, 249117058 Protein Sequence	MSGEPLHVKTPIRDSMALSKMAGTSVYLKMDSAQPSGSFKIRGIGHFCKRWAKQGCAH FVCSSAGNAGMAAAYAARQLGVPATIVVPSTTPALTIERLKNEGATVKVVGELLDEAF ELAKALAKNNPGWVYIPPFDDPLIWEGHASIVKELKETLWEKPGAIALSVGGGGLLCG VVQGLQEVGWDVPIAMETFGAHSFHAATTAGKLVSLPKITSVAKALGVKTVGAQAL KLFQEHPIFSEVISDQEAVAIAIEKFPVDEKILVEPACGAALAAVYSHVIQKLQLEGNL RTPLPSLVIVCGGSNISLAQLRALKEQLGMTNRLPK		
	SEQ ID NO: 267	1031 bp	
NOV16g, 252790334 DNA Sequence	CACCCGTCTCACATGGGACATCATCACCACCATCACATGTCTGGAGAACCCTGCACG TGAAGACCCCATCCGTGACAGCATGGCCCTGTCCAAATGGCCGGCACCAGCGTCTA CCTCAAGATGGACAGTGGCCAGCCCTCCGGCTCCTTCAAGATCCGGGGCATTGGGCAC TTCTGCAAGAGGTGGGCCAAGCAAGGCTGTGCACATTTTGTCTGCTCCTCGCGGGCA ACGCAGGCATGGCGGCTGCATATGCGGCCAGGCAACTCGGCGTCCCCGCCACCATCGT GGTGGCCAGCACCACCTGCTCTCACCATTGAGCGCCTCAAGAATGAAGGTGCCACA GTCAAGGTGGTGGGTGAGTTATTGGATGAAGCCTTCGAGCTGGCCAAAGCCTTAGCGA AGAACAAACCCGGTTGGGTCTACATTCCTCCCTTTGATGACCCCTCATCTGGGAAGG CCACGCTTCCATCGTGAAAGAGCTGAAGGAGACACTGTGGGAAAAGCCGGGGGCCATC GCGCTGTCACTGGGCGGGCGGGGCTGCTGTGTGGAGTGGTCCAGGGGCTGCAGGAGG TGGGCTGGGGGACGTGCCTGTCTCATCGCCATGGAGACTTTTGGTGCCACAGCTTCCA CGCTGCCACCACCGCAGGCAAACTTGTCTCCTGCCCCAAGATCACCAGTGTGCGCAAG GCCCTGGGCGTGAAGACTGTGGGGGCTCAGGCCCTGAAGCTGTTTCAGGAACACCCCA TTTTCTCTGAAGTTATCTCGGACCAGGAGGCTGTGGCCGCCATTGAGAAGTTCGTGGA TGATGAGAAGATCCTGGTGGAGCCCGCTGCGGGGACGCCCTGGCCGCTGTCTATAGC CACGTGATCCAGAAGCTCCAAGTGGAGGGGAATCTCCGAACCCCGCTGCCATCCCTCG TGGTCATCGTCTGCGGGGGCAGCAACATCAGCCTGGCCACAGCTGCGGGCGCTCAAGGA ACAGCTGGGCATGACAAATAGGTTGCCCAAGTGAGCGGCCGCAAG		
	ORF Start: at 1		ORF Stop: TGA at 1018
	SEQ ID NO: 268	339 aa	MW at 35963.4kD
NOV16g, 252790334 Protein Sequence	HPSHMGHHHHHMSGEPLHVKTPIRDSMALSKMAGTSVYLKMDSAQPSGSFKIRGIGH FCKRWAKQGAHFVCSSAGNAGMAAAYAARQLGVPATIVVPSTTPALTIERLKNEGAT VKVVGELLDEAFELAKALAKNNPGWVYIPPFDDPLIWEGHASIVKELKETLWEKPGA IALSVGGGGLLCGVVQGLQEVGWDVPIAMETFGAHSFHAATTAGKLVSLPKITSVAK ALGVKTVGAQALKLFQEHPIFSEVISDQEAVAIAIEKFPVDEKILVEPACGAALAAVY HVIQKLQLEGNLRTPLPSLVIVCGGSNISLAQLRALKEQLGMTNRLPK		
	SEQ ID NO: 269	1036 bp	
NOV16h, 254869149 DNA Sequence	ACATCATCACCACCATCACATGTCTGGAGAACCCTGCACGTGAAGACCCCATCCGT GACAGCATGGCCCTGTCCAAATGGCCGGCACCAGCGTCTACCTCAAGATGGACAGTG CCCAGCCCTCCGGCTCCTTCAAGATCCGGGGCATTGGGCACCTCTGCAAGAGGTGGGC CAAGCAAGGCTGTGCACATTTTGTCTCCTCGGCGGGCAACGCAGGCATGGCGGCT GCATATGCGGCCAGGCAACTCGGCGTCCCCGCCACCATCGTGGTGGCCAGCACCAC CTGCTCTCACCATTGAGCGCCTCAAGAATGAAGGTGCCACAGTCAAGGTGGTGGGTGA GTTATTGGATGAAGCCTTCGAGCTGGCCAAGGCCCTAGCGAAGAACAACCCGGTTGG GTCTACATTCCTCCCTTTGATGACCCCTCATCTGGGAAGGCCACGCTTCCATCGTGA AAGAGCTGAAGGAGACACTGTGGGAAAAGCCGGGGGCCATCGCGTGTCAGTGGGCGG CGGGGGCTGCTGTGTGGAGTGGTCCAGGGGCTGCAGGAGGTGGGCTGGGGGGACGTG CCTGTATCGCCATGGAGACTTTTGGTGCCCAAGCTTCCACGCTGCCACCACCGCAG GCAAACTTGTCTCCTGCCCCAAGATCACCAGTGTGCCAAGGCCCTGGGCGTGAAGAC TGTGGGGGCTCAGGCCCTGAAGCTGTTTCAGGAACACCCCATTTTCTCTGAAGTTATC TCGGACCAGGAGGCTGTGGCCGCCATTGAGAAGTTCGTGGATGATGAGAAGATCCTGG TGGAGCCCGCTGCGGGGCGAGCCCTGGCCGCTGTCTATAGCCACGTGATCCAGAAGCT CCAAGTGGAGGGGAATCTCCGAACCCCGTGCATCCCTCGTGGTCACTGCTGCGGG GGCAGCAACATCAGCCTGGCCAGCTGCGGGCGCTCAAGGAACAGCTGGGCATGACAA		

	ATAGGTTGCCCAAGTGAGCGGCCGCACTCGAGCACCACCACCACCACCAC		
	ORF Start: at 2		ORF Stop: TGA at 1001
	SEQ ID NO: 270	333 aa	MW at 35316.7kD
NOV16h, 254869149 Protein Sequence	HHHHHMSGEPLHVKTPIRDSMALSKMAGTSVYLKMDSAQPSGSFKIRGIGHFCKRWA KQGAHFVCCSAGNAGMAAAYAARQLGVPATIVVPSTTPALTIERLKNEGATVKVVGE LLDEAFELAKALAKNNPGWVYIPFDDPLIWEGHASIVKELKETLWEKPGAIALSVGG GLLCGVVQGLQEVGWGDVPVIAMETFGAHSFHAATTAGKLVSLPKITSAKALGVKT VGAQALKLFQEHPIFSEVISDQEAVAIEKFVDEKILVEPACGAALAAVYSHVIQKL QLEGNLRTPLPSLVVIVCGGSNISLAQLRALKEQLGMTNRLPK		
	SEQ ID NO: 271	988 bp	
NOV16i, CG142631-02 DNA Sequence	CATGATGTCTGGAGAACCCCTGCACGTGAAGACCCCATCCGTGACAGCATGGCCCTG TCCAAAATGGCCGGCACCAGCGTCTACCTCAAGATGGACAGTGCCAGCCCTCCGGCT CCTTCAAGATCCGGGGCATTGGGCACCTCTGCAAGAGGTGGGCCAAGCAAGGCTGTGC ACATTTTGTCTGCTCCTCGGCGGGCAACGCAGGCATGGCGGCTGCATATGCGGCCAGG CAACTCGGCGTCCCGGCCACCATCGTGGTGCCAGCACCACACCTGCTCTCACCATTG AGCGCTCAAGAAATGAAGGTGCCACAGTCAAGGTGGTGGGTGAGTTATTGGATGAAGC CTTCGAGCTGGCCAAGGCCCTAGCGAAGAACAACCCGGGTGGGTCTACATTCCCCC TTTGATGACCCCTCATCTGGGAAGGCCACGCTTCCATCGTGAAAGAGCTGAAGGAGA CACTGTGGGAAAAGCCGGGGGCCATCGCGCTGTCACTGGGCGGCGGGGCTGTCTGTG TGGAGTGGTCCAGGGGCTGCAGGAGGTGGGCTGGGGGACGTGCTGTATCGCCATG GAGACTTTTGGTGCCACAGCTTCCACGCTGCCACCACCGCAAGCAAACTTGTCTCCC TGCCCAAGATCACCAGTGTGTCGAAGGCCCTGGGCGTGAAGACTGTGGGGCTCAGGC CCTGAAGCTGTTTCAGGAACACCCATTCTCTGAAGTTATCTCGGACCAGGAGGCT GTGGCCGCCATTGAGAAGTTCGTGGATGATGAGAAGATCCTGGTGGAGCCCGCTGCG GGGCAGCCCTGGCCGCTGTCTATAGCCACGTGATCCAGAAGCTCCAAGTGGAGGGGAA TCTCCGAACCCCGCTGCCATCCCTCGTGGTTCATCGTCTGCGGGGCGAGCAACATCAGC CTGGCCAGCTGCGGGCGCTCAAGGAACAGCTGGGCATGACAAATAGGTTGCCCAAGT GA		
	ORF Start: ATG at 2		ORF Stop: TGA at 986
	SEQ ID NO: 272	328 aa	MW at 34625.0kD
NOV16i, CG142631-02 Protein Sequence	MMSGEPLHVKTPIRDSMALSKMAGTSVYLKMDSAQPSGSFKIRGIGHFCKRWAKQGA HFVCCSAGNAGMAAAYAARQLGVPATIVVPSTTPALTIERLKNEGATVKVVGELLDEA FELAKALAKNNPGWVYIPFDDPLIWEGHASIVKELKETLWEKPGAIALSVGGGGLLC GVVQGLQEVGWGDVPVIAMETFGAHSFHAATTAGKLVSLPKITSAKALGVKTVGAQA LKLQEHPIFSEVISDQEAVAIEKFVDEKILVEPACGAALAAVYSHVIQKLQLEGN LRTPLPSLVVIVCGGSNISLAQLRALKEQLGMTNRLPK		
	SEQ ID NO: 273	1011 bp	
NOV16j, CG142631-03 DNA Sequence	ACCATGGGACATCATCACCACCATCACATGTCTGGAGAACCCCTGCACGTGAAGACCC CCATCCGTGACAGCATGGCCCTGTCCAAAATGGCCGGCACCAGCGTCTACCTCAAGAT GGACAGTGCCAGCCCTCCGGCTCCTTCAAGATCCGGGGCATTGGGCACCTCTGCAAG AGGTGGGCCAAGCAAGGCTGTGCACATTTGTCTGCTCCTCGGCGGGCAACGCAGGCA TGGCGGCTGCATATGCGGCCAGGCAACTCGGCGTCCCGCCACCATCGTGGTGCCAG CACCACACCTGCTCTCACCATTGAGCGCTCAAGAAATGAAGGTGCCACAGTCAAGGTG GTGGGTGAGTTATTGGATGAAGCCTTCGAGCTGGCCAAGGCCCTAGCGAAGAACAACC CGGGTTGGGTCTACATTTCCCCCTTTGATGACCCCTCATCTGGGAAGGCCACGCTTC CATCGTGAAAGAGCTGAAGGAGACACTGTGGGAAAAGCCGGGGGCCATCGCGCTGTCA GTGGGCGGCGGGGCTGTGTGTGGAGTGGTCCAGGGGCTGCAGGAGGTGGGCTGGG GGGACGTGCTGTATCGCCATGGAGACTTTTGGTGCCACAGCTTCCACGCTGCCAC CACCGCAGGCAAACTTGTCTCCCTGCCAAGATCACCAGTGTGCAAGGCCCTGGGC GTGAAGACTGTGGGGGCTCAGGCCCTGAAGCTGTTTCAGGAACACCCATTCTCTG AAGTTATCTCGGACCAGGAGGCTGTGGCCGCCATTGAGAAGTTCGTGGATGATGAGAA GATCCTGGTGAGCCCGCTGCGGGGAGCCCTGGCCGCTGTCTATAGCCACGTGATC CAGAAGCTCCAAGTGGAGGGGAATCTCCGAACCCCGCTGCCATCCCTCGTGGTCACTCG TCTGCGGGGCGAGCAACATCAGCCTGGCCAGCTGCGGGCGCTCAAGGAACAGCTGGG CATGACAAATAGGTTGCCCAAGTGA		
	ORF Start: at 1		ORF Stop: TGA at 1009

	SEQ ID NO: 274	336 aa	MW at 35606.0kD
NOV16j, CG142631-03 Protein Sequence	TMGHHHHHMSGEPLHVKTPIRDSMALSKMAGTSVYLMDSAQPSGSFKIRGIGHFCK RWAKQGCAHFVCSSAGNAGMAAAYAARQLGVPATIVVPSTTPALTIERLKNEGATVKV VGELLDEAFELAKALAKNNPGWVYIPPFDDPLIWEGHASIVKELKETLWEKPGAIALS VGGGGLLCGVVQGLQEVGWGDPVIAMETFGAHSFHAATTAGKLVSLPKITSVAKALG VKTVGAQALKLFQEHPIFSEVISDQEAVAAIEKFVDDEKILVEPACGAALAAVYSHVI QKLQLEGNLRTPPLSLVVIVCGGSNISLAQLRALKEQLGMTNRLPK		
	SEQ ID NO: 275	1008 bp	
NOV16k, CG142631-04 DNA Sequence	ACCATGATGTCTGGAGAACCCCTGCACGTGAAGACCCCATCCGTGACAGCATGGCCC TGTCCAAAATGGCCGGCACCAGCGTCTACCTCAAGATGGACAGTGGCCAGCCCTCCGG CTCCTTCAAGATCCGGGGCATTGGGCACTTCTGCAAGAGTGGGCCAAGCAAGGCTGT GCACATTTGTCTGCTCCTCGCGGGCAACGCAGGCATGGCGGCTGCATATCGGCCA GGCAACTCGGCGTCCCCGCCACCATCGTGGTGGCCAGCACCACCTGCTCTCACCAT TGAGCGCTCAAGAATGAAGGTGCCACAGTCAAGGTGGTGGGTGAGTTATTGGATGAA GCCTTCGAGCTGGCCAAGGCCCTAGCGAAGAACAACCCGGGTGGGTCTACATTCCCC CCTTTGATGACCCCTCATCTGGGAAGGCCACGCTTCCATCGTGAAAGAGCTGAAGGA GACACTGTGGGAAAAGCCGGGGGCCATCGCGCTGTCACTGGGCGCGGGGGCCTGTGT TGTGGAGTGGTCCAGGGGCTGCAGGAGGTGGGCTGGGGGACGTGCCTGTATCGCCA TGGAGACTTTTGGTGGCCACAGCTTCCACGCTGCCACCACCGCAGGCAAACTGTCTC CCTGCCCAAGATCACCAGTGTGTGCCAAGGCCCTGGGCGTGAAGACTGTGGGGGCTCAG GCCCTGAAGCTGTTTCAGGAACACCCATTTTCTCTGAAGTTATCTCGGACCAGGAGG CTGTGGCCGCCATTGAGAAGTTTCGTGGATGATGAGAAGATCCTGGTGGAGCCCGCCTG CGGGGCAGCCCTGGCGCTGTCTATAGCCACGTGATCCAGAAGCTCAACTGGAGGGG AATCTCCGAACCCCGCTGCCATCCCTCGTGGTCATCGTCTGCGGGGGCAGCAACATCA GCCTGGCCAGCTGCGGGCGCTCAAGGAACAGCTGGGCATGACAAATAGTTGCCCAA GCATCATCACCACCATCACTGA		
	ORF Start: at 1		ORF Stop: TGA at 1006
	SEQ ID NO: 276	335 aa	MW at 35549.0kD
NOV16k, CG142631-04 Protein Sequence	TMSGEPLHVKTPIRDSMALSKMAGTSVYLMDSAQPSGSFKIRGIGHFCKRWAKQGC AHFVCSSAGNAGMAAAYAARQLGVPATIVVPSTTPALTIERLKNEGATVKVVGELLDE AFELAKALAKNNPGWVYIPPFDDPLIWEGHASIVKELKETLWEKPGAIALSVGGGGLL CGVVQGLQEVGWGDPVIAMETFGAHSFHAATTAGKLVSLPKITSVAKALGVKTVGAQ ALKLFQEHPIFSEVISDQEAVAAIEKFVDDEKILVEPACGAALAAVYSHVIQKLQLEG NLRTPPLSLVVIVCGGSNISLAQLRALKEQLGMTNRLPKHHHHHH		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 16B.

Table 16B. Comparison of NOV16a against NOV16b through NOV16k.		
Protein Sequence	NOV16a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV16b	1..328	328/328 (100%)
	1..328	328/328 (100%)
NOV16c	1..328	323/328 (98%)
	2..329	324/328 (98%)
NOV16d	1..328	323/328 (98%)
	1..328	324/328 (98%)
NOV16e	2..328	322/327 (98%)
	1..327	323/327 (98%)
NOV16f	2..328	322/327 (98%)

	1..327	323/327 (98%)
NOV16g	2..328	322/327 (98%)
	13..339	323/327 (98%)
NOV16h	2..328	322/327 (98%)
	7..333	323/327 (98%)
NOV16i	1..328	323/328 (98%)
	1..328	324/328 (98%)
NOV16j	2..328	322/327 (98%)
	10..336	323/327 (98%)
NOV16k	1..328	323/328 (98%)
	2..329	324/328 (98%)

Further analysis of the NOV16a protein yielded the following properties shown in Table 16C.

Table 16C. Protein Sequence Properties NOV16a	
PSort analysis:	0.8500 probability located in endoplasmic reticulum (membrane); 0.4400 probability located in plasma membrane; 0.1000 probability located in mitochondrial inner membrane; 0.1000 probability located in Golgi body
SignalP analysis:	No Known Signal Sequence Predicted

5 A search of the NOV16a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 16D.

Table 16D. Geneseq Results for NOV16a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV16a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAU23764	Novel human enzyme polypeptide #850 - Homo sapiens, 340 aa. [WO200155301-A2, 02-AUG-2001]	5..321 23..338	192/317 (60%) 246/317 (77%)	e-106
ABB89752	Human polypeptide SEQ ID NO 2128 - Homo sapiens, 329 aa. [WO200190304-A2, 29-NOV-2001]	5..321 12..327	192/317 (60%) 246/317 (77%)	e-106
AAM40622	Human polypeptide SEQ ID NO 5553 - Homo sapiens, 340 aa. [WO200153312-A1, 26-JUL-2001]	5..321 23..338	192/317 (60%) 246/317 (77%)	e-106

AAM38836	Human polypeptide SEQ ID NO 1981 - Homo sapiens, 329 aa. [WO200153312-A1, 26-JUL-2001]	5..321 12..327	192/317 (60%) 246/317 (77%)	e-106
AAU23238	Novel human enzyme polypeptide #324 - Homo sapiens, 340 aa. [WO200155301-A2, 02-AUG-2001]	5..321 23..338	192/317 (60%) 246/317 (77%)	e-106

In a BLAST search of public sequence databases, the NOV16a protein was found to have homology to the proteins shown in the BLASTP data in Table 16E.

Table 16E. Public BLASTP Results for NOV16a				
Protein Accession Number	Protein/Organism/Length	NOV16a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
P20132	L-serine dehydratase (EC 4.2.1.13) (L-serine deaminase) - Homo sapiens (Human), 328 aa.	1..328 1..328	328/328 (100%) 328/328 (100%)	0.0
Q8VBT2	Similar to serine dehydratase - Mus musculus (Mouse), 327 aa.	1..328 1..327	270/328 (82%) 294/328 (89%)	e-151
DWRTT	L-serine dehydratase (EC 4.2.1.13) - rat, 327 aa.	1..326 1..326	269/326 (82%) 289/326 (88%)	e-151
Q91X68	Similar to serine dehydratase - Mus musculus (Mouse), 313 aa.	1..313 1..313	260/313 (83%) 281/313 (89%)	e-147
Q8WW81	Hypothetical 23.0 kDa protein - Homo sapiens (Human), 218 aa.	1..217 1..217	214/217 (98%) 214/217 (98%)	e-122

PFam analysis predicts that the NOV16a protein contains the domains shown in the Table 16F.

Table 16F. Domain Analysis of NOV16a			
Pfam Domain	NOV16a Match Region	Identities/ Similarities for the Matched Region	Expect Value
PALP	4..298	97/378 (26%) 221/378 (58%)	3.8e-64

## 5 Example 17.

The NOV17 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 17A.

Table 17A. NOV17 Sequence Analysis			
	SEQ ID NO: 277	1146 bp	
NOV17a, CG151359-01 DNA Sequence	ATGAGTTGGACTGTACCTGTTGTGTGGGCCAGCCAGAGAGTGAGCTCGGCAGGAGCGA ATTTTCTGTGCCTGGGGATGGCCCTGTGTCCCCGTGAGGCAGCGTGCATGCCACTCAT GGGCACCTGGCTCTTCACCTCCGTGAGCAAGATGGCGACTGTGAAGAGTGAGCTTATT GAGTGCTTCACTTCCGAGGAGCCCTTTCATCACAGAAAGGTCTCCATCACAGGAAGTG GATCAGTGGGCATGGCCTGCGCTACCAGCATCTTATTAAGGCTTGAGTGATGAAGT TGCCTTTGTGGATCTTGATGAAGGCAAACTGAAAGGTGAGACAATGGATCTTCAACAT GACAGCCCTTTCATGAAAATGTCAAATATTGTTTGTAGCAAGATTACCTTGTCCACG CAAACCCCATCTAGTGATTATCACAGCAGGTGCACGCCGAGAAAAGGGAGAAATGCG CTTTAATTTAGTCCCGCAAAATGTGGCCATCTTCAAGTTAATGATTTCAGTATTGTC CAGCAGAGCCCCCTCTGCAAATAATTATTGTTTCCAATCCAGTAGATATCTTAACTT ACGTAGCCTGGAAGTTGAGTGCATTTCCAAAAACCGTGTATTGGAAGCGGCTGTAA TCTGGATACTGTTCTTTCAATTCTTCATTGGACAAAAGCTTGGTATCCACTCTGAA AGCTGCCGTGGATGGATCCTCGGAGAGCATGGAGACTCAAGTGTTCTGTGTGGAGTG GAATGAACATAGCTGGTGTCTTTTGAAGGATCTGAACTCTGATATAGGAAGTATGATA AGATCCTGAGAAATGAAAAATGTCCACAAAGAAGTGATTGCTAGTGCCTATGAGATT ATTGAAATGAAAAGTTCTACTTCGTGGGCCATTGGCCTATCTGGAGCTGATTTAACAG AAAGTATTTTGAAGAATCTTAGGAGAAAAACATCCAGTTTCCACCATAATTAAGGGCCT CTACGGAATAAATGAAGAAGTCTTCTCAGTATTCCTTCTTTGTTGGAGAGAAGGGT ATTACCAACCTTATAAAGAGAAAAGCTGACCCCTGAAGAGGAGGCCCATCTGAAAAAGA GTGCAAAAACACTTTGGGAAATTCAGAAGGAGCTTGAGACTTAA		
	ORF Start: ATG at 1		ORF Stop: TAA at 1144
	SEQ ID NO: 278	381 aa	MW at 42104.6kD
NOV17a, CG151359-01 Protein Sequence	MSWTVPVVWASQRVSSAGANFLCLGMALCPRQAACMPLMGTLWFTSVSKMATVKSELI ECFTSEEPFHHRKVSITGTGSVGMACATSILLKGLSDELAFLVLDDEGLKGETMDLQH DSPFMKMSNIVCSKDYLVNANPHLVIIITAGARREKGMRFNLVRQNVAFKLMISIV QQSPLCKLIIVSNPVDILTIVVANKLSAFPKNRIVGSGCNLDTVRFQFFIGQKLGIHSE SCRGWILGEHGDSSVPVWSGMNIAGVLLKDLNSDIGTDKDPKWKNVHKEVIASAYEI IEMKSSTSWAIGLSGADLTESILKNLRRKHPVSTIIKGLYGINEEVFLSIPSLFGEKG ITNLIKRLTPREEAHLKKSATLWEIQKELET		

Further analysis of the NOV17a protein yielded the following properties shown in Table 17B.

Table 17B. Protein Sequence Properties NOV17a	
PSort analysis:	0.6736 probability located in nucleus; 0.5701 probability located in mitochondrial matrix space; 0.3952 probability located in microbody (peroxisome); 0.2847 probability located in mitochondrial inner membrane
SignalP analysis:	Cleavage site between residues 49 and 50

- 5 A search of the NOV17a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 17C.

Table 17C. Geneseq Results for NOV17a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV17a Residues/ Match	Identities/ Similarities for the Matched	Expect Value

		<b>Residues</b>	<b>Region</b>	
AAU11432	Human testicular lactate dehydrogenase A - Homo sapiens, 381 aa. [CN1313342-A, 19-SEP-2001]	1..380 1..380	328/380 (86%) 344/380 (90%)	0.0
AAG89135	Human secreted protein, SEQ ID NO: 255 - Homo sapiens, 381 aa. [WO200142451-A2, 14-JUN-2001]	1..380 1..380	328/380 (86%) 344/380 (90%)	0.0
AAY36058	Extended human secreted protein sequence, SEQ ID NO. 443 - Homo sapiens, 381 aa. [WO9931236-A2, 24-JUN-1999]	1..380 1..380	321/380 (84%) 336/380 (87%)	0.0
AAM42058	Human polypeptide SEQ ID NO 6989 - Homo sapiens, 372 aa. [WO200153312-A1, 26-JUL-2001]	44..380 35..371	221/337 (65%) 271/337 (79%)	e-128
AAM40272	Human polypeptide SEQ ID NO 3417 - Homo sapiens, 332 aa. [WO200153312-A1, 26-JUL-2001]	50..380 1..331	218/331 (65%) 268/331 (80%)	e-127

In a BLAST search of public sequence databases, the NOV17a protein was found to have homology to the proteins shown in the BLASTP data in Table 17D.

<b>Table 17D. Public BLASTP Results for NOV17a</b>				
<b>Protein Accession Number</b>	<b>Protein/Organism/Length</b>	<b>NOV17a Residues/ Match Residues</b>	<b>Identities/ Similarities for the Matched Portion</b>	<b>Expect Value</b>
Q9BYZ2	L-lactate dehydrogenase A-like (EC 1.1.1.27) - Homo sapiens (Human), 381 aa.	1..380 1..380	328/380 (86%) 344/380 (90%)	0.0
Q96LI2	CDNA FLJ25463 fis, clone TST09242 (Lactate dehydrogenase A-like) - Homo sapiens (Human), 381 aa.	1..380 1..380	325/380 (85%) 342/380 (89%)	0.0
DEM5LM	L-lactate dehydrogenase (EC 1.1.1.27) chain M - mouse, 332 aa.	50..380 1..331	220/331 (66%) 271/331 (81%)	e-129
P06151	L-lactate dehydrogenase A chain (EC 1.1.1.27) (LDH-A) (LDH muscle subunit) (LDH-M) - Mus musculus (Mouse), 331 aa.	51..380 1..330	219/330 (66%) 270/330 (81%)	e-128
Q9XT87	L-lactate dehydrogenase A chain (EC 1.1.1.27) (LDH-A) (LDH muscle subunit) (LDH-M) - Monodelphis	52..380 2..330	219/329 (66%) 269/329 (81%)	e-127

	domestica (Short-tailed grey opossum), 331 aa.			
--	--	--	--	--

PFam analysis predicts that the NOV17a protein contains the domains shown in the Table 17E.

Table 17E. Domain Analysis of NOV17a			
Pfam Domain	NOV17a Match Region	Identities/ Similarities for the Matched Region	Expect Value
ldh	67..210	63/156 (40%) 120/156 (77%)	9.1e-55
ldh_C	212..380	68/179 (38%) 148/179 (83%)	4.4e-67

### Example 18.

5 The NOV18 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 18A.

Table 18A. NOV18 Sequence Analysis			
	SEQ ID NO: 279	1015 bp	
NOV18a, CG152227-01 DNA Sequence	CTCTGCTGCTTTAGTTTCGGAGTGTGTTGGCGACGGGGCAGCGCAGATGTGGAGGCTC ATGTCGAGGTTTAAATGCATTCAAAGGACTAATACCATACTGCACCATTGAGAATGT CCAAGCACACAGATGCAGCAGAAGAGGTGCTATTGGAAAAAAGGTTGCGCGGGAGT CATAACACTAAACAGACCAAAGTTCCTCAATGCACTGACTCTTAATATGATTGGCGAG ATTTATCCACAGCTAAAGAAGTGGGAACAAGATCCTGAAACTTTCCTGATCATTATAA AGGGAGCAGGAGGAAAGGCTTCTGTGCGGGGGTGATATCAGAGTGATCTCGGAAGC TGAAGAGGCAAAACAGAAGATAGCTCCAGTTTCTTCAGAGAAGAATATATGCTGAAT AATGCTGTTGGTTCTTGCCAGAAACCTTATGTGCACTTATTCATGGAATTACAATGG GTGGGGGAGTTGGTCTCTCAGTCCATGGGCAATTCGAGTGGCTACAGAAAAGTGTCT TTTTGCTATGCCAGAACTGCAATAGGACTGTTCCCTGATGTGGGTGGAGGTTATTTCT TTGCCACGACTCCAAGGAAACTTGGTTACTTCCTTGCAATTAACAGGATTACAGACTAA AAGGAAGAGATGTGTACAGAGCAGGAATGTCTACACACTTTGTAGATTCTGAAAAGTT GGCCATGTTAGAGGAAGATTTGTTAGCCTTGAAATCTCCTTCAAAGAAAATATTGCA TCTGTCTTAGAAAATTACCATACAGAGTCTAAGATTGATCGAGACAAGTCTTTTATAC TTGAAGACCAGAGTCCAAAATGGAAACCAGCTGATCTAAAAGAAGCTACTGAGGAAGA TTTGAATAATCACTTTAAGTCTTTGGGAAGCAGTGATTGAAATTTTGGGTGACAGG CTTTAAAGGTATATTTGTAGCATGGGTTGGCAATCTACAGCATGTGGGCCAAATCCA GCCTGCTGCCTGTTTTATATACCCTGTA		
	ORF Start: ATG at 47.		ORF Stop: TGA at 917
	SEQ ID NO: 280	290 aa	MW at 32497.3kD
NOV18a, CG152227-01 Protein Sequence	MWRLMSRFNAFKRTNTILHHLRMSKHTDAEEVLLLEKKGCAVITLNRPKFLNALTIN MIRQIYPQLKKWEQDPETFLIIKAGGKAFKAGGDIRVISEAEKAKQKIAPVFFREE YMLNNAVGSQKPYVALIHGITMGGGVGLSVHGQFRVATEKCLFAMPETAIGLFPDVG GGYFLPRLQGLGYFLALTGPRLLKGRDVYRAGIATHFVDSKLALEEDLLALKSPSK ENIASVLENYHTESKIDRKSFILEDQSPKWKPADLKEATEEDLNHFKSLGSSDLKF		
	SEQ ID NO: 281	1311 bp	
NOV18b,	AGTCCGGGAGATTCTCGCTCTGCTGCTTTAGTTTCGGAGTGTGTTGGCGACGGGGCAGC		

CG152227-02 DNA Sequence	GCGAGATGTGGAGGCTCATGTCGAGGTTTAAATGCATTCAAAGGACTAATACCATACT GCACCATTGAGAATGTCCAAGCACACAGATGCAGCAGAAGAGGTGCTATTGGAAAAA AAAGGTTGCGCGGGAGTCATAACACTAAACAGACCAAAGTTCTCTCAATGCACTGACTC TTAATATGATTCCGCAGATTTATCCACAGCTAAAGAAAGTGGGAACAAGATCCTGAAAC TTTCGTGATCATTATAAAGGGAGCAGGAGGAAAGGCTTTCTGTGCCGGGGGTGATATC AGAGTGATCTCGGAAGCTGAAAAGGCAAAACAGAAGATAGTCCAGTTTCTTCAGAG AAGAATATATGCTGAATAATGCTGTTGGTTC'TTGCCAGAAACCTTATGTTGCACCTTAT TCATGGAATTACAATGGGTGGGGGAGTTGGTCTCTCAGTCCATGGGCAATTTGAGAGTG GCTACAGAAAAGTGTCTTTTGTCTATGCCAGAACTGCAATAGGACTGTTCCCTGATG TGGGTGGAGGTTATTTCTTTGCCACGACTCCAAGGAAAACCTTGGTTACTTCCTTGCAT TAACGGATTACAGCTAAAAGGAAGAGATGTGTACAGAGCAGGAATTGCTACACACTTT GTAGATTCTGAAAAGTTGGCCATGTTAGAGGAAGATTTGTTAGCCTTGAAATCTCCTT CAAAAGAAAATATTGCATCTGTCTTAGAAAATTACCATACAGAGTCTAAGATTGATCG AGACAAGTCTTTTATAC'TTGAGGAACACATGGACAAAATAAACAGTTGTTTTTCAGCC AATACTGTGGAAGAAATTTATTGAAAACCTTACAGCAAGATGGTTCATCTTTTGCCCTAG AGCAATTGAAGGTAATTAATAAAATGTCTCCAACATCTCTAAAGATCACACTAAGGCA ACTCATGGAGGGGTCTTCAAAGACCTTGCAAGAAGTACTAATATGGAGTATCGGCTA AGTCAAGCTTGTATGAGAGGTCATGACTTTTCATGAAGGCGTTAGAGCTGTTTTAATTG ATAAAGACCAGAGTCCAAAATGGAACCAGCTGATCTAAAAGAAGTTACTGAGGAAGA TTTGAATAATCACTTTAAGTCTTTGGGAAGCAGTGATTTGAAATTTTGAGGTGACAGG CTTTTAAGGTATATTTTGTAGCATGGGTGGCAATCTACAGCATGTGGGCCAAATCCA GCCTGCTGCCTGTTTTTATATACCCTGTAAGCAAG		
	ORF Start: ATG at 64		ORF Stop: TGA at 1207
	SEQ ID NO: 282	381 aa	MW at 42907.1kD
NOV18b, CG152227-02 Protein Sequence	MWRLMSRFNAFKRTNTIILHHLRMSKHTDAABEVLLLEKKGCAGVITLNRPKFLNALTLN MIRQIYPQLKKWEQDPETFVIIIGAGGKAFCAAGDIRVISEAEKAKQKIAPVFFREE YMLNNAVSGCQKPYVALIHGIMGGVGVLVSHGQFRVATEKCLFAMPETAIGLFPDVG GGYFFATTPRKTWLLPCINGFRLKGRDVYRAGIATHFVDESKLAMLEDDLALKSPSK ENIASVLENYHTESKIDRDKSFILLEHMDKINSFCSANTVEEIIENLQQDGSSFALEQ LKVINKMSPSTSLKITLRQLMEGSSKTLQEVLTMEYRLSQACMRGHDHFEGVRAVLIDK DQSPKWKPADLKEVTEEDLNNHFKSLGSSDLKF		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 18B.

Table 18B. Comparison of NOV18a against NOV18b.		
Protein Sequence	NOV18a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV18b	1..278	246/278 (88%)
	1..278	250/278 (89%)

Further analysis of the NOV18a protein yielded the following properties shown in Table 18C.

5

Table 18C. Protein Sequence Properties NOV18a	
PSort analysis:	0.6784 probability located in mitochondrial matrix space; 0.3893 probability located in microbody (peroxisome); 0.3672 probability located in mitochondrial inner membrane; 0.3672 probability located in mitochondrial intermembrane space
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV18a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 18D.

Table 18D. Geneseq Results for NOV18a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV18a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAW81135	Human 3-hydroxyisobutyryl-coenzyme A hydrolase - Homo sapiens, 381. aa. [WO9851782-A2, 19-NOV-1998]	1..278 1..278	259/278 (93%) 261/278 (93%)	e-147
AAG75795	Human colon cancer antigen protein SEQ ID NO:6559 - Homo sapiens, 178 aa. [WO200122920-A2, 05-APR-2001]	2..176 1..175	158/175 (90%) 159/175 (90%)	1e-86
ABB61217	Drosophila melanogaster polypeptide SEQ ID NO. 10443 - Drosophila melanogaster, 351 aa. [WO200171042-A2, 27-SEP-2001]	29..278 8..250	131/253 (51%) 171/253 (66%)	2e-63
AAG23865	Arabidopsis thaliana protein fragment SEQ ID NO: 27329 - Arabidopsis thaliana, 378 aa. [EP1033405-A2, 06-SEP-2000]	23..254 1..232	98/233 (42%) 148/233 (63%)	9e-50
AAG23866	Arabidopsis thaliana protein fragment SEQ ID NO: 27330 - Arabidopsis thaliana, 374 aa. [EP1033405-A2, 06-SEP-2000]	32..254 6..228	97/224 (43%) 145/224 (64%)	1e-49

5 In a BLAST search of public sequence databases, the NOV18a protein was found to have homology to the proteins shown in the BLASTP data in Table 18E.

Table 18E. Public BLASTP Results for NOV18a				
Protein Accession Number	Protein/Organism/Length	NOV18a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q9BS94	Similar to 3-hydroxyisobutyryl-coenzyme A hydrolase - Homo sapiens (Human), 333 aa.	1..278 1..278	261/278 (93%) 263/278 (93%)	e-148
Q92931	3-hydroxyisobutyryl-coenzyme A hydrolase - Homo sapiens (Human), 381. aa.	1..278 1..278	246/278 (88%) 250/278 (89%)	e-138

Q8QZS1	Similar to 3-hydroxyisobutyryl-coenzyme A hydrolase - Mus musculus (Mouse), 385 aa.	2..278 7..282	207/277 (74%) 238/277 (85%)	e-118
Q9VF79	CG5044 protein - Drosophila melanogaster (Fruit fly), 351 aa.	29..278 8..250	131/253 (51%) 171/253 (66%)	6e-63
Q960K8	LD47223p - Drosophila melanogaster (Fruit fly), 385 aa.	29..278 42..284	131/253 (51%) 171/253 (66%)	6e-63

PFam analysis predicts that the NOV18a protein contains the domains shown in the Table 18F.

Table 18F. Domain Analysis of NOV18a			
Pfam Domain	NOV18a Match Region	Identities/ Similarities for the Matched Region	Expect Value
ECH	42..213	54/176 (31%) 112/176 (64%)	2.3e-17

#### Example 19.

The NOV19 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 19A.

5

Table 19A. NOV19 Sequence Analysis			
	SEQ ID NO: 283	1935 bp	
NOV19a, CG152392-01 DNA Sequence	GTGCTGGCTTGCCCTGCAAATTGTGTCTGCAGCAAGACTGAGATCAATTGCCGGCGGC CGGACGATGGGAACCTCTTCCCCCTCCTGGAAGGGCAGGATTCAGGGAAACAGCAATGG GAACGCCAGTATCAACATCACGGACATCTCAAGGAATATCACTTCCATACACATAGAG AACTGGCGCAGTCTTACACGCTCAACGCCGTGGACATGGAGCTCTACACCGGACTTC AAAAGCTGACCATCAAGAACTCAGGACTTCGGAGCATTACGCCAGAGCCTTTGCCAA GAACCCCATTTGCGTTATATAAACCTGTCAAGTAACCGGCTCACCACACTCTCGTGG CAGCTCTTCCAGACGCTGAGTCTTCGGGAATTGCAGTTGGAGCAGAACTTTTCAACT GCAGCTGTGACATCCGCTGGATGCAGCTCTGGCAGGAGCAGGGGGAGGCCAAGCTCAA CAGCCAGAACCTCTACTGCATCAATGCTGATGGCTCCCAGCTTCTCTCTTCCGCATG AACATCAGTCAGTGTGACCTTCTGAGATCAGCGTGAGCCACGTCAACCTGACCGTAC GAGAGGGTGACAATGCTGTTATCACTTGCAATGGCTCTGGATCACCCCTTCTGATGT GGACTGGATAGTCACTGGGCTGCAGTCCATCAACACTCACCAGACCAATCTGAACTGG ACCAATGTTTCATGCCATCAACTTGACGCTGGTGAATGTGACGAGTGAGGACAATGGCT TCACCTGACGTGCATTGCAGAGAACGTGGTGGGCATGAGCAATGCCAGTGTGTCCT CACTGTCTACTATCCCCACGTGTGGTGAGCCTGGAGGAGCCTGAGCTGCGCCTGGAG CACTGCATCGAGTTGTGGTGGTGGCAACCCCCACCAACGCTGCAGCTGGCTGCACA ATGGGCAGCCTCTGCGGGAGTCCAAGATCATCCATGTGGAATACTACCAAGAGGGAGA GATTTCGAGGGGCTCCTGCTCTTCAACAAGCCCACCCACTACAACAATGGCAACTAT ACCCTCATTGCCAAAACCCACTGGGACAGCCAAACGACCATCAATGGCCACTTCC TCAAGGAGCCCTTTCAGTTGACGAAGTGAGTCCCACACCTCCTATCACTGTGACCCA CAAACCAGAAGAAGACACTTTTGGGGTATCCATAGCAGTTGGAGTGTGCTTTTGCC TGTGTCCTGTTGGTGGTGTCTTCGTCATGATCAACAAATATGGTCGACGGTCCAAAT TTGGAATGAAGGGTCCCGTGGCTGTCACTAGTGGTGAGGAGGACTCAGCCAGCCCACT GCACCACATCAACCACGGCATCACCACGCCCTCGTCACTGGATGCGGGGCCGACACT GTGGTCATTGGCATGACTCGCATCCCTGTCAATTGAGAACCCCACTTCCGTCAGG		

	GACACAACCTGCCACAAGCCGGACACGTGGGTCTTTTCAAACATAGACAATCATGGGAT ATTAAACTTGAAGGACAATAGAGATCATCTAGTCCCATCAACTCACTATATATATGAG GAACCTGAGGTCCAGAGTGGGGAAGTGTCTTACCCAAGGTCACATGGTTTCAGAGAAA TTATGTTGAATCCAATAAGCCTTCCCGGACATTCCAAGCCTCTTAACCATGGCATCTA TGTTGAGGATGTCAATGTTTATTTTCAGCAAAGGACGTCATGGCTTTTAAAACTCCTT TTAAGCCTCCTTGTGTTTGATGTCACCTTGGTAGGCTGGGCCCTCTGAGAGGTTGGAAG CTCTAGGCATTGTTCTCTTTGGATCCAGGGATGCTAAGTAGAACTGCATGAGCCACC AGTGCCCCGGCACCTTTAACACCACCAGATGGGTGTTTTCCCCCATCCACCACTGGC AGGGCTTGCCAGGAGTAAGAG		
	ORF Start: at 1		ORF Stop: TAA at 1729
	SEQ ID NO: 284	576 aa	MW at 64294.1kD
NOV19a, CG152392-01 Protein Sequence	VLACPANCVCCKTEINCRRPDDGNLFPLLEGQDSGNSNGNASINITDISRNITSIHIE NWRSLHTLNAVDMELYTGLOKLTIKNSGLRSIQPRAFAKNPHLYINLSSNRLTTLWS QLFQTLRLRLQLEQNFFNCSCDIRWMLWQEQGEAKLNSQNYLCINADGSQPLFRM NISQCDLPEISVSHVNLTVREGDNAVITCNGSGSPLPDVDWI VTGLQSINTHTNLNW TNVHAINLTLVNVTSEDNGFTLTCAENVVGMSNASVALTVYPPRVVSLEEPELRLE HCIEFVVRGNPPPTLHWHNGQPLRESKIIHVEYYQEGEISEGCLLFNKPHTYNNNGNY TLIAKNPLGTANQTINGHFLKEFPVDEVSPPTITVTHKPEEDTFGVSTAVGLAFAFA CVLLVVVFVMINKYGRRSKFGMKGPVAVISGEEDSASPLHHINHGITTPSSLDAGPDT VVIGMTRIPVIENTPQYFRQGHNCHKPDTWVFSNIDNHGILNLKDNRDHLPSTHYIYE EPEVQSGEVSYPRSHGFREIMLNPISLPGHKSPLNHGIYVEDVNVYFSKGRHGF		

Further analysis of the NOV19a protein yielded the following properties shown in Table 19B.

Table 19B. Protein Sequence Properties NOV19a	
PSort analysis:	0.8357 probability located in mitochondrial inner membrane; 0.8200 probability located in plasma membrane; 0.3000 probability located in microbody (peroxisome); 0.2000 probability located in endoplasmic reticulum (membrane)
SignalP analysis:	No Known Signal Sequence Predicted

- 5 A search of the NOV19a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 19C.

Table 19C. Geneseq Results for NOV19a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV19a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAAY51602	Human truncated trkC receptor protein - Homo sapiens, 612 aa. [US6027927-A, 22-FEB-2000]	1..576 29..612	573/584 (98%) 575/584 (98%)	0.0
AAR81627	Human trkC receptor protein mutant - Homo sapiens, 830 aa. [WO9525795-A1, 28-SEP-1995]	1..494 29..521	490/494 (99%) 493/494 (99%)	0.0

AAY06595	Neurotrophin-3 receptor TrkC - Homo sapiens, 825 aa. [WO9940103-A1, 12-AUG-1999]	1..494 29..530	491/502 (97%) 493/502 (97%)	0.0
AAM50853	Human receptor tyrosine kinase TrkC - Homo sapiens, 839 aa. [WO200203071-A2, 10-JAN-2002]	1..494 29..530	490/502 (97%) 493/502 (97%)	0.0
AAY51601	Human trkC receptor protein - Homo sapiens, 839 aa. [US6027927-A, 22-FEB-2000]	1..494 29..530	490/502 (97%) 493/502 (97%)	0.0

In a BLAST search of public sequence databases, the NOV19a protein was found to have homology to the proteins shown in the BLASTP data in Table 19D.

Table 19D. Public BLASTP Results for NOV19a				
Protein Accession Number	Protein/Organism/Length	NOV19a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q96CY4	Hypothetical 68.5 kDa protein - Homo sapiens (Human), 612 aa.	1..576 29..612	574/584 (98%) 575/584 (98%)	0.0
I73633	gene trkC protein - human, 612 aa.	1..576 29..612	573/584 (98%) 575/584 (98%)	0.0
Q9Z2P9	Neurotrophin-3 receptor non-catalytic isoform 2 - Mus musculus (Mouse), 612 aa.	1..576 29..612	553/584 (94%) 568/584 (96%)	0.0
A55178	neurotrophin receptor trkC precursor - human, 825 aa.	1..494 29..530	491/502 (97%) 493/502 (97%)	0.0
O75682	TRKC protein - Homo sapiens (Human), 839 aa.	1..494 29..530	491/502 (97%) 493/502 (97%)	0.0

PFam analysis predicts that the NOV19a protein contains the domains shown in the Table 19E.

Table 19E. Domain Analysis of NOV19a			
Pfam Domain	NOV19a Match Region	Identities/ Similarities for the Matched Region	Expect Value
LRRNT	3..30	9/31 (29%) 23/31 (74%)	0.00013
LRR	100..123	8/25 (32%) 22/25 (88%)	0.0043

LRRCT	132..180	13/54 (24%) 40/54 (74%)	2.4e-10
ig	196..258	20/65 (31%) 43/65 (66%)	4.8e-07

**Example 20.**

The NOV20 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 20A.

Table 20A. NOV20 Sequence Analysis			
	SEQ ID NO: 285	1201 bp	
NOV20a, CG152453-01 DNA Sequence	GCCCTTCTGGCAGGAAGAGGAAGATGTCGTGTGCTCAGGCGGATGATGCGGGTTTCCAA TCGCTCTCTCCTCGCCTTCATCTTCTTCTCCCTCTCTTCGTCTGTCTGTACTTC ATCTATGTGGCCCCAGGCATCGCCAACACATATCTCTTTATGGTACAAGCTCGAGGTA TAATGTTGAGAGAAAATGTGAAAACAATAGGTCATATGATCAGGCTGTACACAAATAA AAACAGTACGCTCAACGGTACAGATTATCCCGAAGGCAATAATTCAAGTGATTATCTT GTTCAAACAACAACGTATCTCCCGGAAAACCTCACATACTCACCATACTCCCTGTCT CAGAAAAGCTGCCTTATATGCGAGGATTCTCAATGTCAATGTAAGCGAAGTCAGTTT TGATGAAATTTCATCAACTCTTCTCCAAGGATTAGATATTGAGCCAGGGGGTCATTGG AGGCCAAAAGACTGTAAACCCAGATGGAAGGTGGCAGTTCTCATTCTTTCCGTAATC GCCATGAACATCTTCCAATTTTTTTCTTACATCTGATTCCAATGCTCCAGAAGCAGCG GCTGGAATTTGCGTTTTATGTCTATTGAACAGACTGGCACACAACCTTTTAACCGTGCG ATGCTTTTCAATGTGGGCTTCAAAGAGGCCATGAAAGACAGTGTCTGGGACTGTGTAA TCTTCCACGATGTGGATCATCTACCTGAAAATGACCGGAACATTACCGATGTGGAGA AATGCCACGTCATTTGCTGCAAAGCTGGATAAATACATGTATATTCTCCATATAAA GAATTTTTTGGTGGTGAAGTGGGCTGACAGTGAACAATTTAGAAAGATCAATGGTT TTCCTAATGCCTTCTGGGGATGGGAGGAGAAGATGATGACCTTTGGAACAGAGTTCA CTATGCTGGATATAATGTAACCAGACCAGAGGGAGACTTAGGAAAATACAAGTCAATT CCTCATCACCATAGAGGTGAAGTCCAGTTTTTAGGACGGTATAAATTACTAAGGTATT CCAAGGAGCGTCAGTACATCGATGGACTGAACAATTTAATATATAGGCCAAAAAATACT GGTTGATAGGTTGTATACAAACATATCTGTAAACCTCATGCCAGAGTTAGCTCCAATC GAAGACTATTAAAGAAGTGGCTGTCTGTTGGCAAGGTAGACC		
	ORF Start: ATG at 24		ORF Stop: TAA at 1170
	SEQ ID NO: 286	382 aa	MW at 44913.2kD
NOV20a, CG152453-01 Protein Sequence	MSVLRRMRVSNRSLAFI FFFSLSSCLYFIYVAPGIANTYLFMVQARGIMLRENVK TIGHMIRLYTNKNSTLNGTDYPEGNSSDYLVQTTTYLPENFTYSPYLPCEKLPYMR GFLNVNVSEVSPDEIHQLFSKDLIDIEPGGHWPKDCKPRWKVAVLIPFRNRHEHLP FLHLIPMLQKQRLFAFYVIEQTGTQPFNRAMLFNVGFKEAMKDSVWDCVIFHDVDHL PENDRNYYGCGEMPRHFAAKLDKMYILPYKEFFGGVSGLTVEQFRKINGFPNFWGW GGEDDDLWNRVHYAGYNVTRPEGLGKYKSI PHHHRGEVQFLGRYKLLRYSKERQYID GLNNLIYRPKILVDRLYTNISVNLMPELAPIEDY		
	SEQ ID NO: 287	1062 bp	
NOV20b, CG152453-03 DNA Sequence	GATGTCGTGTGCTCAGGCGGATGATGCGGGTTTCCAATCGCTCTCTCTCGCCTTCATC TTCTTCTTCTCCCTCTCTTCGTCTGTCTGTACTTCATCTATGTGGCCCCAGGCATCG ATTATCCCGAAGGCAATAATTCAAGTGATTATCTTGTTCAAACAACAACGTATCTCCC GGAAAACCTCACATACTCACCATACCTCCCTGTCCAGAAAAGCTGCCTTATATGCGA GGATTCTCAATGTCAATGTAAGCGAAGTCAGTTTTGATGAAATTCATCAACTCTTCT CCAAGGATTTAGATATTGAGCCAGGGGGTCATTGGAGGCCAAAAGACTGTAAACCCAG ATGGAAGGTGGCAGTTCTCATTCTTTCCGTAATCGCCATGAACATCTTCCAATTTTT TTCTTACATCTGATTCCAATGCTCCAGAAGCAGCGGCTGGAATTTGCGTTTTATGTCA TTGAACAGACTGGCACACAACCTTTTAACCGTGCGATGCTTTTCAATGTGGGCTTCAA AGAGGCCATGAAAGACAGTGTCTGGGACTGTGTAATCTTCCACGATGTGGATCATCTA		

	CCTGAAAATGACCGGAAC TATTACGGATGTGGAGAAATGCCACGTCATTTTGCTGCAA AGCTGGATAAATACATGTATATCTTCCATATAAAGAATTTTTGGTGGTGTAAGTGG GCTGACAGTGAACAATTTAGAAAGATCAATGGTTTTCCATATGCCTTCTGGGGATGG GGAGGAGAAGATGATGACCTTTGGAACAGAGTTCACATATGCTGGATATAATGTAACCA GACCAGAGGGAGACTTAGGAAAATACAAGTCAATTCCTCATCACCATAGAGGTGAAGT CCAGTTTTTAGGACGGTATAAATTACTAAGGTATTCGAAGGACGTCAGTACATCGAT GGACTGAACAATTTAATATATAGGCCAAAAATACTGGTTGATAGGTTGTATACAAACA TATCTGTAAACCTCATGCCAGAGTTAGCTCCAATCGAAGACTATTAAAGAAGTGGCT GTCGTGGCAAGGTAGACC		
	ORF Start: ATG at 2		ORF Stop: TAA at 1031
	SEQ ID NO: 288	343 aa	MW at 40460.0kD
NOV20b, CG152453-03 Protein Sequence	MSVLRRMMRVSNRSLAFIFFFSLSSSSCLYFIYVAPGIDYPEGNSSDYLVTQTTTYLP ENFTYSPYLPCEKLPYMRGFLNVNVSEVSFDEIHQLFSKDLIDIEPGGHWPKDCKPR WKVAVLIPFRNRHEHLPIFFLHLIPMLQQRLEFAFYVIEQTGTQPFNRAMLFNVGFK EAMKDSVWDCVIFHDVDHLPENDRNYYGCGEMPRHFAAKLDKMYILPYKEFFGGVSG LTVEQFRKINGFPNAPFWGWWGGEEDDLWNRVHYAGYNVTRPEGDLGKYKSI PHHHRGEV QFLGRYKLLRYSKERQYIDGLNNLIYRPKILVDRLYTNISVNLMP ELAPIEDY		
	SEQ ID NO: 289	1100 bp	
NOV20c, CG152453-02 DNA Sequence	ATGTCCTGTGCTCAGCGGATGATGCGGGTTTCCAATCGCTCTCTCCTCGCCTTCATCT TCTTCTCTCCCTCTCTTCGTCCTGTCTGTACTTCATCTATGTGGCCCCAGGCATCGC CAACACACATCTCTTTATGGTACAAGCTCGAGGTATAATGTTGAGAGAAAATGTGAAA ACAATAGGTCATATGATCAGGCTGTACACAATAAAAACAGTACGCTCAACGGTACAG ATTATCCCGAAGGCAATAATTCAAGTGATTATCTTGTTCAAACAACAACGTATCTCCC GGAAAACCTTACATACTCACCATACCTCCCCTGTCCAGAAAAGCTGCCTTATATGCGA GGATTCCTCAATGTCAATGTAAGCGAAGTCAGTTTTGATGAAATTCATCAACTCTTCT CCAAGGATTTAGATATTGAGCCAGGGGTCATTGGAGGCCAAAAGACTGTAAACCCAG ATGGAAGAAGCAGCGCTGGAATTTGCGTTTATGTCAATGAACAGACTGGCACACAA CCTTTTAACCGTGCATGCTTTTCAATGTGGGCTTCAAAGAGGCCATGAAAGACAGTG TCTGGGACTGTGTAATCTTCCACGATGTGGATCATCTACCTGAAAATGACCGGAAC TA TTACGGATGTGGAGAAATGCCACGTCATTTTGCTGCAAAGCTGGATAAATACATGTAT ATTCTTCCATATAAAGAATTTTTGGTGGTGTAAAGTGGGCTGACAGTGAACAATTTA GAAAGATCAATGGTTTTCTAATGCCTTCTGCGGATGGGGAGGAGAAGATGATGACCT TTGGAACAGAGTTCACTATGCTGGATATAATGTAACCAGACCAGAGGGAGACTTAGGA AAATACAAGTCAATTCCTCATCACCATAGAGGTGAAGTCCAGTTTTTAGGACGGTATA AATTACTAAGGTATTTCAAGGAGCGTCAGTACATCGATGGACTGAACATTTAATATA TAGGCCAAAAATACTGGTTGATAGGTTGTATACAAACATATCTGTAAACCTCATGCCA GAGTTAGCTCCAATCGAAGACTATTAAAGAAGTGGCTGTGCTGGCAAGGTAGACC		
	ORF Start: ATG at 1		ORF Stop: TAA at 1069
	SEQ ID NO: 290	356 aa	MW at 41753.4kD
NOV20c, CG152453-02 Protein Sequence	MSVLRRMMRVSNRSLAFIFFFSLSSSSCLYFIYVAPGIANTHLFMVQARGIMLRENVK TIGHMIRLYTNKNSTLNGTDYPEGNSSDYLVTQTTTYLPENFTYSPYLPCEKLPYMR GFLNVNVSEVSFDEIHQLFSKDLIDIEPGGHWPKDCKPRWKKQRLEFAFYVIEQTGTQ PFNRAMLFNVGFK EAMKDSVWDCVIFHDVDHLPENDRNYYGCGEMPRHFAAKLDKMY ILPYKEFFGGVSGLTVEQFRKINGFPNAPFWGWWGGEEDDLWNRVHYAGYNVTRPEGDLG KYKSI PHHHRGEVQFLGRYKLLRYSKERQYIDGLNNLIYRPKILVDRLYTNISVNLMP ELAPIEDY		

- 5 Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 20B.

Table 20B. Comparison of NOV20a against NOV20b and NOV20c.		
Protein Sequence	NOV20a Residues/ Match Residues	Identities/ Similarities for the Matched Region

NOV20b	1..382	343/382 (89%)
	1..343	343/382 (89%)
NOV20c	1..382	355/382 (92%)
	1..356	356/382 (92%)

Further analysis of the NOV20a protein yielded the following properties shown in Table 20C.

Table 20C. Protein Sequence Properties NOV20a	
PSort analysis:	0.8541 probability located in lysosome (lumen); 0.7189 probability located in outside; 0.2757 probability located in microbody (peroxisome); 0.1000 probability located in endoplasmic reticulum (membrane)
SignalP analysis:	Cleavage site between residues 28 and 29

5 A search of the NOV20a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 20D.

Table 20D. Geneseq Results for NOV20a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV20a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAW81569	Human lactosyl ceramide synthase - Homo sapiens, 382 aa. [JP10295371-A, 10-NOV-1998]	1..382 1..382	382/382 (100%) 382/382 (100%)	0.0
ABG23077	Novel human diagnostic protein #23068 - Homo sapiens, 404 aa. [WO200175067-A2, 11-OCT- 2001]	1..382 23..404	381/382 (99%) 382/382 (99%)	0.0
AAW81567	Rat lactosyl ceramide synthase - Rattus sp, 382 aa. [JP10295371-A, 10-NOV-1998]	1..382 1..382	360/382 (94%) 376/382 (98%)	0.0
AAW81568	Mouse lactosyl ceramide synthase - Mus sp, 382 aa. [JP10295371-A, 10-NOV-1998]	1..382 1..382	362/382 (94%) 374/382 (97%)	0.0
AAB26791	Human galactoside transferase I- type homologous protein - Homo sapiens, 343 aa. [CN1257925-A, 28-JUN-2000]	1..382 1..343	342/382 (89%) 343/382 (89%)	0.0

In a BLAST search of public sequence databases, the NOV20a protein was found to have homology to the proteins shown in the BLASTP data in Table 20E.

<b>Table 20E. Public BLASTP Results for NOV20a</b>				
<b>Protein Accession Number</b>	<b>Protein/Organism/Length</b>	<b>NOV20a Residues/ Match Residues</b>	<b>Identities/ Similarities for the Matched Portion</b>	<b>Expect Value</b>
Q9UBX8	Beta-1,4-galactosyltransferase 6 (EC 2.4.1.-) (Beta-1,4-GalTase 6) (Beta4Gal-T6) (b4Gal-T6) (UDP-galactose:beta-N-acetylglucosamine beta-1,4-galactosyltransferase 6) (UDP-Gal:beta-GlcNAc beta-1,4-galactosyltransferase 6) [Includes: Lactosylceramide synthase (EC 2.4.1.-) (LacCer synthase) (UDP-Gal:glucosylceramide beta-1,4-galactosyltransferase)] - Homo sapiens (Human), 382 aa.	1..382 1..382	382/382 (100%) 382/382 (100%)	0.0
O88419	Beta-1,4-galactosyltransferase 6 (EC 2.4.1.-) (Beta-1,4-GalTase 6) (Beta4Gal-T6) (b4Gal-T6) (UDP-galactose:beta-N-acetylglucosamine beta-1,4-galactosyltransferase 6) (UDP-Gal:beta-GlcNAc beta-1,4-galactosyltransferase 6) [Includes: Lactosylceramide synthase (EC 2.4.1.-) (LacCer synthase) (UDP-Gal:glucosylceramide beta-1,4-galactosyltransferase)] - Rattus norvegicus (Rat), 382 aa.	1..382 1..382	360/382 (94%) 376/382 (98%)	0.0
Q9WVK5	Beta-1,4-galactosyltransferase 6 (EC 2.4.1.-) (Beta-1,4-GalTase 6) (Beta4Gal-T6) (b4Gal-T6) (UDP-galactose:beta-N-acetylglucosamine beta-1,4-galactosyltransferase 6) (UDP-Gal:beta-GlcNAc beta-1,4-galactosyltransferase 6) [Includes: Lactosylceramide synthase (EC 2.4.1.-) (LacCer synthase) (UDP-Gal:glucosylceramide beta-1,4-galactosyltransferase)] - Mus musculus (Mouse), 382 aa.	1..382 1..382	362/382 (94%) 374/382 (97%)	0.0
Q8WZ95	Beta-1,4-galactosyltransferase - Homo sapiens (Human), 343 aa.	1..382 1..343	342/382 (89%) 343/382 (89%)	0.0

O43286	Beta-1,4-galactosyltransferase 5 (EC 2.4.1.-) (Beta-1,4-GalTase 5) (Beta4Gal-T5) (b4Gal-T5) (UDP-galactose:beta-N-acetylglucosamine beta-1,4-galactosyltransferase 5) (UDP-Gal:beta-GlcNAc beta-1,4-galactosyltransferase 5) (EC 2.4.1.-) (Beta-1,4-GalT II) - Homo sapiens (Human), 388 aa.	1..382 1..388	273/388 (70%) 321/388 (82%)	e-169
--------	--	------------------	--------------------------------	-------

PFam analysis predicts that the NOV20a protein contains the domains shown in the Table 20F.

Table 20F. Domain Analysis of NOV20a			
Pfam Domain	NOV20a Match Region	Identities/ Similarities for the Matched Region	Expect Value
Galactosyl_T_2	108..375	157/329 (48%) 266/329 (81%)	3.2e-187

### Example 21.

The NOV21 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 21A.

Table 21A. NOV21 Sequence Analysis			
	SEQ ID NO: 291	1327 bp	
NOV21a, CG152547-01 DNA Sequence	ATGGGCCGCTACTCTGGCAAGACGTGCCGGCTGCTCTTCATGCTGGTGCTCACCCTCG CCTTCTTCGTGGCGGAGCTGGTCTCCGGCTACCTGGGCAACTCCATCGCGCTGCTCTC CGACTCCTTCAACATGCTCTCCGACCTGATCTCGCTGTGCGTGGGCGCTGAGCGCCGGC TACATCGCCCGCGCCCCACCCGGGGCTTCAGCGCCACCTACGGCTACGCCCCGCGCCG AGGTGGTGGGCGCGCTGAGCAACGCGGTCTTCTCACCAGCGCTCTGCTTCAACATCTT CGTGGAGGCCGTGCTGCGCCTGGCCCGGCCGAGCGCATCGATGACCCGAGCTGGTG CTCATCGTCCGCGTCTTGGGGCTGTTGGTCAACGTGGTGGGGCTGCTCATCTTCCATC ACCAATCCCTAATCTCAAGTAATCAGGGACACAAACACTGCGGAAGGCCGAGGGTCC TCTGCCTAGGAAAACAGAAACACCCAGAATGAGCCAGAAGACATGATGAAAAAGAG AAAAAGTCTGAAGCTCTGAATATCAGAGGTGTACTTTTGCATGTGATGGGAGATGCC TGGGGTCCGTGGTTGTGGTTCATCACGGCCATCATATTCTATGTGCTTCCCTGAAGAG TGAGGACCCGTGTAAGTGGCAGTGTTACATTGACCCAGCCTGACTGTCTCATGGTC ATCATCATTTTGTCTATCTGCCCTTCCCGCTTATCAAGGAGACCGCTGCCATTCTGTCTAC AGATGGTCCCAAAGGAGTCAACATGGAAGAGCTGATGAGTAACTCTCTGCTGTGCC TGGAATTAGCAGTGATACATGAAGTGCACATCTGGGAACCTGTAAGTGGAAAGATTATT GCCACCCTGCACATCAAGTATCCTAAGGACAGGGGATATCAAGATGCCAGCACAAAAA TTCGAGAAATCTCCACCATGCGGGAATCCACAATGTGACCATCCAGTTTGAATGT GGACTTGAAGGAACCCCTGGAGCAGAAGGACTTACTGTGCTCTGCAACTACCCCTGC ATCTCCAAGGGCTGTGCTAGCAGCTGTGTTGTCCCCCGGGGCACTGCCTCTGGCTC ACGTCAATGGCTGTGCTGAGCAATGGTGGGCCCTCTCTAGACACATACGGAAGTGA TGGCCTCAGTAGAAGAGACGCAAGAGAAGTGGCTATTGAAGTGTCTTTGATAGCTGT CTGAGTGACCACGGACAATGTCTTAACAAAACCTCAGGAGGACCAATGTTATGTCAACA GAACGCATTTTAAATCTGGTACTCACATAATCAGACCATATAGACGAGAAG		

	ORF Start: ATG at 1		ORF Stop: TAA at 1288
	SEQ ID NO: 292	429 aa	MW at 46990.2kD
NOV21a, CG152547-01 Protein Sequence	MGRYSGKTCRLFLMLVLTVAFFVAELVSGYLGNSIALLSDFNMLSDLISLCVGLSAG YIARRPTRGFSATYGYARAEVVGALSNAVFLTALCFTIFVEAVLRLARPERIDDP ELVLIVGVLGLLVNVVGLLI FHHQSLISSNQGHKHCGRPQGPLPRKTRNTQNEP EDMMKKEKKSEALNIRGVLLHVMGDALGSVVVVITAIIFYVLPLKSEDP CNWQCYIDPSLTVLMV I I I LSSAFPLIKETAAILLQMPKGVNMEELMSKLS AVPGI SSVHEVHIWELVSGKIIATLHIKYPKDRGYQDASTKIREIFHHAGI HNVTIQFENVDLKEPLEQKDLLLLCNSPCISKGCAQLCCPPGALPLAHVNG CAEHNGGPSLDTYGSDGLSRRDAREVAIEVSLDSC LSDHGQCLNK TQEDQCYVNRTHF		

Further analysis of the NOV21a protein yielded the following properties shown in Table 21B.

Table 21B. Protein Sequence Properties NOV21a	
PSort analysis:	0.6400 probability located in plasma membrane; 0.4600 probability located in Golgi body; 0.3700 probability located in endoplasmic reticulum (membrane); 0.1000 probability located in endoplasmic reticulum (lumen)
SignalP analysis:	Cleavage site between residues 30 and 31

- 5 A search of the NOV21a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 21C.

Table 21C. Geneseq Results for NOV21a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV21a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
ABP51303	Human MDDT SEQ ID NO 325 - Homo sapiens, 520 aa. [WO200240715-A2, 23-MAY-2002]	1..429 36..520	410/485 (84%) 413/485 (84%)	0.0
AAU99906	Human 83378 metal transporter protein - Homo sapiens, 485 aa. [WO200240656-A2, 23-MAY-2002]	1..429 1..485	408/485 (84%) 411/485 (84%)	0.0
AAM52621	Human zinc ion transport protein 26 - Homo sapiens, 240 aa. [WO200181539-A2, 01-NOV-2001]	190..429 1..240	238/240 (99%) 238/240 (99%)	e-138
AAG66785	Zinc transporter homologue ZnT-1-22 - Homo sapiens, 199 aa. [WO200171000-A1, 27-SEP-2001]	231..429 1..199	197/199 (98%) 197/199 (98%)	e-112

AAU69449	Human purified secretory polypeptide #18 - Homo sapiens, 349 aa. [WO200162918-A2, 30-AUG-2001]	1..290 36..349	240/346 (69%) 243/346 (69%)	e-111
----------	--	-------------------	--------------------------------	-------

In a BLAST search of public sequence databases, the NOV21a protein was found to have homology to the proteins shown in the BLASTP data in Table 21D.

Table 21D. Public BLASTP Results for NOV21a				
Protein Accession Number	Protein/Organism/Length	NOV21a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q9NPW0	Hypothetical 26.3 kDa protein - Homo sapiens (Human), 240 aa.	190..429 1..240	239/240 (99%) 239/240 (99%)	e-138
Q9Y6M5	Zinc transporter 1 (ZnT-1) - Homo sapiens (Human), 507 aa.	1..398 1..485	181/493 (36%) 249/493 (49%)	2e-72
Q9VZR4	CG17723 protein (LD22804P) - Drosophila melanogaster (Fruit fly), 449 aa.	1..359 1..378	148/390 (37%) 228/390 (57%)	5e-68
Q06808	Oxidative stress resistance - Saccharomyces cerevisiae (Baker's yeast), 429 aa.	5..351 3..398	143/402 (35%) 222/402 (54%)	6e-61
P20107	Zinc/cadmium resistance protein - Saccharomyces cerevisiae (Baker's yeast), 442 aa.	5..351 3..398	143/402 (35%) 222/402 (54%)	6e-61

PFam analysis predicts that the NOV21a protein contains the domains shown in the Table 21E.

Table 21E. Domain Analysis of NOV21a			
Pfam Domain	NOV21a Match Region	Identities/ Similarities for the Matched Region	Expect Value
Cation_efflux	11..333	101/358 (28%) 259/358 (72%)	2.2e-68

## 5. Example 22.

The NOV22 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 22A.

Table 22A. NOV22 Sequence Analysis
------------------------------------

	SEQ ID NO: 293	1047 bp	
NOV22a, CG152646-01 DNA Sequence	AGGCTGGGCACGAGGACCATGCTGGGCCGAGCCTCCGAGAAGTTTCTGCGGCACTGA AACCAAGGCCAAATTACACCAACAGAGCTCTGTCAAAAATGTCTCTCTTATCAAGAA GACCAAGTTTCTAAATGCCTACATTACTGTGTGTCAGAAGAGGTGGCCTTAAACAAGCT GAAGAATCAGAAAAGAGATATAAGAATGGACAGTCACTTGGGGATTAGATGGAATTC CTATTGCAGTAAAAGACAATTTTCAGCACTTCTGGCATTGAGACAACATGTGCATCAAA TATGCTGAAAGGTTATATACCACCTTATAATGCTACAGTAGTTTCAGAAGTTGTTGGAT CAGGGAGCTCTACTAATGGGAAAAACAAATTTAGATGAGTTTGCTATGGGATCTGGGA GCACAGATGGTGTATTTGGACCAGTTAAAAACCCCTGGAGTTATTCAAACAATATGG TCACAGATGTGACATTGATTTGTCCACTGAAGCCATGTATGCTGCAACCAGACGAGAA GGGTTTAAATGATGTGGTGAGAGGAAGAATTCCTCAGGAACTTTTTCTTATTAAGAG AAACTATGAAAATTATTTGTCAAAGCACAGAAAGTGAGACGCCTCATTGCTAATGA CTTTGTAAATGCTTTTAACTCTGGAGTAGATGTCTTGCTAACTCCCACCACCTTGAGT GAGGCAGTACCATACTTGGAGTTCATCAAAGAGGACAACAGAACCCGAAGTGCCCAGG ATGATATTTTACACAAGCTGTAAATATGGCAGGATTGCCAGCAGTGAGTATCCCTGT TGCACCTCAAACCAGGGGTTGCCAATAGGACTGCAGTTTATGGACGTGCGTTTTGT GACCAGCAGCTTCTACAGTAGCCAAATGGTTGAAAAACAAGTACAGTTTCTCTGTTA TTCAACTTCAAGAACTCATGGATGATTGTTTCAGCAGTCCTTGAAAATGAAAAGTTAGC CTCTGTCTCTCTAAACAGTAAACATATCTTACAAATTAAATGACTTTTAGGCTGGG TGC		
	ORF Start: ATG at 19		ORF Stop: TAA at 1006
	SEQ ID NO: 294	329 aa	MW at 36411.3kD
NOV22a, CG152646-01 Protein Sequence	MLGRSLREVSAALKQGGITPTELCQKCLSLIKTKFLNAYITVSEEVALKQAESEKR YKNGQSLGDLDGIPIAVKDNFSTSGIETTCASNMLKGYIPPNATVQKLLDQGALLM GKTNLDEFAMGSGSTDGVFGPVKNPWSYSKQYGHRCIDIDLSTEAMYAATRRREGFNDVV RGRILSGNFFLLKENYENYFVKAQKVRRLIANDFVNAFNSGVDVLLTPTTLSEAVPYL EFIKEDNRTRSAQDDIFTQAVNMAGLPAVSI PVALSNQGLPIGLQFIGRAFCDDQLLT VAKWFEKQVQFPVIQLQELMDDCSAVLENEKLASVSLKQ		

Further analysis of the NOV22a protein yielded the following properties shown in Table 22B.

Table 22B. Protein Sequence Properties NOV22a	
PSort analysis:	0.6500 probability located in cytoplasm; 0.1000 probability located in mitochondrial matrix space; 0.1000 probability located in lysosome (lumen); 0.0000 probability located in endoplasmic reticulum (membrane)
SignalP analysis:	No Known Signal Sequence Predicted

- 5 A search of the NOV22a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 22C.

Table 22C. Geneseq Results for NOV22a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV22a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
ABP41274	Human ovarian antigen HOSED43, SEQ ID NO:2406 - Homo sapiens.	147..329 81..263	182/183 (99%) 183/183 (99%)	e-100

	263 aa. [WO200200677-A1, 03-JAN-2002]			
ABB05695	Human nucleic acid management protein clone fbr2_78c12 - Homo sapiens, 528 aa. [WO200198454-A2, 27-DEC-2001]	147..329 346..528	182/183 (99%) 183/183 (99%)	e-100
AAE18112	Human glutamyl-tRNA (Gln) amidotransferase-like enzyme - Homo sapiens, 528 aa. [WO200200703-A2, 03-JAN-2002]	147..329 346..528	182/183 (99%) 183/183 (99%)	e-100
AAU19422	Human diagnostic and therapeutic polypeptide (DITHP) #8 - Homo sapiens, 549 aa. [WO200162927-A2, 30-AUG-2001]	147..329 367..549	182/183 (99%) 183/183 (99%)	e-100
AAB94654	Human protein sequence SEQ ID NO:15566 - Homo sapiens, 528 aa. [EP1074617-A2, 07-FEB-2001]	147..329 346..528	182/183 (99%) 183/183 (99%)	e-100

In a BLAST search of public sequence databases, the NOV22a protein was found to have homology to the proteins shown in the BLASTP data in Table 22D.

Table 22D. Public BLASTP Results for NOV22a				
Protein Accession Number	Protein/Organism/Length	NOV22a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q9NV19	Hypothetical 57.5 kDa protein (Similar to hypothetical protein FLJ10989) - Homo sapiens (Human), 528 aa.	147..329 346..528	182/183 (99%) 183/183 (99%)	e-100
Q9H0R6	Hypothetical 57.5 kDa protein - Homo sapiens (Human), 528 aa.	147..329 346..528	182/183 (99%) 183/183 (99%)	e-100
Q9CZN8	2700038P16Rik protein - Mus musculus (Mouse), 525 aa.	147..329 342..524	163/183 (89%) 169/183 (92%)	6e-88
Q9HA60	CDNA FLJ12189 fis, clone MAMMA1000841, moderately similar to putative amidase (EC 3.5.1.4) - Homo sapiens (Human), 303 aa.	1..148 1..148	148/148 (100%) 148/148 (100%)	4e-80
Q9VE09	GATA protein - Drosophila melanogaster (Fruit fly), 508 aa.	147..305 336..499	89/164 (54%) 114/164 (69%)	6e-43

Pfam analysis predicts that the NOV22a protein contains the domains shown in the Table 22E.

Table 22E. Domain Analysis of NOV22a			
Pfam Domain	NOV22a Match Region	Identities/ Similarities for the Matched Region	Expect Value
Amidase	22..142	58/126 (46%) 98/126 (78%)	1.5e-41
Amidase	148..289	62/170 (36%) 114/170 (67%)	7.6e-35

### Example 23.

5 The NOV23 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 23A.

Table 23A. NOV23 Sequence Analysis			
	SEQ ID NO: 295	1935 bp	
NOV23a, CG152959-01 DNA Sequence	AGAGGGCTCAAGAGGGGCAGCCCCGCATAGAGGAGATGCGAGCTCTGCGCTCTGCCAG GGCCCCGAGCCCGTCAGAGGGCCCGCCGCGCCCGGGAAGCCACCGCGGCCCCCTC ACTCCTAGAGGAAGGGAGCACCGCGAGGCTCACGGCAGGGCCCTGGCGCCGGGCAGGG CGAGCCTCGGAAGCCGCTGGAGGACGTGCTGTGGCTGCAGGAGGTCTCCAACCTGTC AGAGTGGCTGAGTCCCAGCCCTGGGCCCTGAGCCGGGTCCCTTCCGCAAGCGCCAC CGATCCGAGGCTGCGGGCAGCCGTTATCCCGTGGTTTAAAGCTGCCGCGCGCTC ACCAAGTCTCTTCCGCGTCTGCTTCCGCGTCGGGCCCCGGCGGGGCGGGCGGGCGG TGGAGCCCGCGCCGGCCTGACGTCAACACACCTCCCTGGGACTGCGTCACTGGTGC GCGCCGCGGGTCAGGGCGCAATGGCGCGCTGGGCGGGATGGGCTGCGACTGTGTC GGTGTGCGGGCCGAGCGGCCCGAGTCGCGCGGCTGGGCGGCTGGGCCCCGGG CTGTGCTGCTGGGTGTCACTGTTCTCCTGCCTCAGCCTCGCCTGCTCTACATGGGCA GCCTCTACGTCTGGAAGAGCGAACTGCCCAGGGACCATCCGCGGTCAAGCGACG CTTACCAGCGTCTGGTGGTGTCCAGTCTCTCACCCCTGTGCGTGTGCTCTGGAGG GAACACAGGCATCCAGGCACATCCCTGCTCACCTGATGGGCTTCAGGCTGGAGGG CATTTTCCAGCGCGCTGCTGCCCTGTTGCTGACCATGATTCTTTTCTGGGCCCA CTGATGCAGCTCTCTATGGATTGCCCTTGTGACCTGGCAGATGGGCTGAAGGTTGTCC TGGCCCCCGCTCCTGGGCCCGCTGCCTCACAGACATGCGTTGGCTGCGGAACCAAGT GATCGCCCCGCTGACAGAGGAGCTGGTGTTCGGGGCTGTATGCTGCCCATGTTAGCA CCGTGCATGGGCTGGGCCCTGCTGTGTTACCTGCCCCGCTCTTTTGGAGTTGCCC ATTTTACCATAATTATTGAGCAGCTGCGTTTCCGCCAGAGCAGCGTGGGAACATCTT CTTGTCTGCTGCGTTCCAGTTCCCCTACACAGCTGTCTCGGTGCCTACACTGCTTTC CTCTTCATCCGCACAGGACACCTGATTGGGCCGTTCTCTGCCATTCTTCTGCAATT ACATGGGTTTCCAGCTGTTTGGCGGGCTTGGAGACCCACAGAGGCGGCCCTGCT GGCAGGCTATGCCCTGGGTGTGGGACTCTTCTGCTTCTGCTCCAGCCCTCACGGAC CCCAAGCTCTACGGCAGCCTTCCCCTTTGTGTGCTTTTGGAGCGGCAGGGGACTCAG AGGCTCCCCTGTCTCCTGACCTATGCTCCTGGATACGCTATGAACCTCACCCGCTC CCCAGCCCTCCCCACCAAGGGTACTGCAGGGGAAGGGCTGGCTGGGCTCCCCGAGAT CTCAGGAATTTTGTAGGGGATTGAAGCCAGAGCTAGTTGCGTCCCAGGGACCAAGAG AAAGAAGCAGATATCCAAAGGTGCAGCCCTTTTGAAGGGGTGTTTACGAGCAGCT GTGAGTGAGGGGACAAGGGGAGGTCCCAGGAGCCACACTCCCTTCTCACTTTGG ACTGTGCTTCTTCTAGCTCCTCTGCCTCTGAAAAGCTGCTCGGGGTTTTTATTAT AAAACCTCTCCCCACCCCCCACTTCTGGGTTTTCTCATGCTTTTGTG ATCAGTACTTTGTATTGGGATATTAAAGAGATTTAACTTGGGTAAAAA AAAAAAAAAAAAAAAAAAAA		

	ORF Start: ATG at 485		ORF Stop: TGA at 791
	SEQ ID NO: 296	102 aa	MW at 10925.7kD
NOV23a, CG152959-01 Protein Sequence	MAALGGDGLRLLSVSRPERPPESAALGGLGPGLCWVSFVFSCLSLACSYMGSlyVWKS ELPRDHPAVIKRRFTSVLVVSSLSPLCVLLWRELtGIQAHPCSP		
	SEQ ID NO: 297	1472 bp	
NOV23b, CG152959-02 DNA Sequence	GTCACtGGTGCgCGCCGCGGGTCAGGGCGCAATGGCGGCGCTGGGCGGGGATGGGCTG CGACTGCTGTGCGGTGTGCGGGCCGGAGCGGCCGCCGAGTCGCGGGCGCTGGGCGGGCC TGGGCCCCGGGCTGTGCTGCTGGGTGTCAGTGTCTCCTGCCTCAGCCTCGCCTGCTC CTACGTGGGcAGCCTCTACGTCTGGAAGAGCGAACTGCCcAGGGACCATCCCGCGGTc ATCAAGCGACGCTTCACCAGCGTCCTGGTGGTGTCCAGTCTCTACCCCTGTGCGTGC TGCTCTGGAGGGAACtCACAGGCATCCAGCCAGGCACATCCCTGCTCACCCTGATGGG CTTCAGGCTGGAGGGCATTtTCCAGCGGCGCTGCTGCCCCGTtTGCTGACCATGATT CTTTTCCTGGGCCCCACTGATGCAGCTCTCTATGGATTGCCCTTGtGACCTGGCAGATG GGCTGAAGGTtTGCTGGCCCCCGCTCCTGGGCGCGCTGCCTCACAGACATGCGTTG GCTGCGGAACCAAGTGATCGCCCCGCTGACAGAGGAGCTGGTGTtCCGGGCGCTGTATG CTGCCCATGTtAGCACCGTGCATGGGCGCTGGGCGCTGCTGTGTtCACCTGCCCGCTCT TTTTTGAGTtGCCCATtTTCACCATATTATTGAGCAGCTGCGTtTCCGCCAGAGCAG CGTGGGGAACATCTTCTGTCTGCTGCGTtCCAGTtTCTCTACACAGCTGTCTTCGGT GCCTACACTGCTtTCTCTTCATCCGCACAGGACACCTGATTGGGCGGTTCTCTGCG ATTCTTCTGCAATTACATGGGTtTCCAGCTGTTTGCGCGGCGCTTGGAGCACCCACA GAGGCGGCCCTGCTGGCAGGCTATGCCCTGGGTGTGGGACTCTTCTGCTTCTGCTC CAGCCCCTCACGGACCCCAAGCTCTACGGCAGCCTTCCCTTtGTGTGCTtTTGGAGC GGGCAGGGGACTCAGAGGCTCCCTGTGCTCCTGACCTATGCTCCTGGATACGCTATG AACTCTACCGGCTCCCCAGCCCTCCCCACCAAGGGGTACTGCAGGGGAAGGGCTGGC TGGGGTCCCCGAGATCTCAGGAATtTTTGtAGGGGATTGAAGCCAGAGCTAGTTGCGT CCCAGGGACCAAGAGAAAGAAGCAGATATCCAAAGGGTGCAGCCCCtTTTGAAAGGGG TGtTTACGAGCAGCTGTGAGTGAAGGGGACAAGGGGcAGGTCCCAGGACCCACACACTC CCTTCTCACTtTTGGACTGCTGCTTCTCTTAGCTCCTCTGCCTCTGAAAAGCTGCTCG GGGTTTTTTATTTATAAAACCTCTCCCCACCCCCACCCCCAAACTTCTGGGTTTT CTCATTGTCTTTTGCATCAGTACTTTGTATTGGGATATTAAAGAGATTAACTTGGG TAAAAAAAAAAAAAAAAAAAAA		
	ORF Start: ATG at 32		ORF Stop: TGA at 1019
	SEQ ID NO: 298	329 aa	MW at 35832.2kD
NOV23b, CG152959-02 Protein Sequence	MAALGGDGLRLLSVSRPERPPESAALGGLGPGLCWVSFVFSCLSLACSYVGSlyVWKS ELPRDHPAVIKRRFTSVLVVSSLSPLCVLLWRELtGIQPGTSLLTLMGFRLEGIFPAA LLPLLLTMILFLGPLMQLSMDPCDLADGLKVVLAPRWARCLTDMRWLRNQVIAPLT EELVFRACMLPMLAPCMGLGPAVFTCPFFGVAHFHHIIEQLRFRQSSVGNIFLSAAF QFSYTAVFGAYTAFLFIRTGHLIGPVLCHSFCNYMGFPAVCAALEHPQRRLLAGYAL GVGLFLLLLQPLTDPKLYGSLPLCVLLERAGDSEAPLCS		

- 5 Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 23B.

Table 23B. Comparison of NOV23a against NOV23b.		
Protein Sequence	NOV23a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV23b	1..96	95/96 (98%)
	1..96	96/96 (99%)

Further analysis of the NOV23a protein yielded the following properties shown in Table 23C.

<b>Table 23C. Protein Sequence Properties NOV23a</b>	
<b>PSort analysis:</b>	0.7000 probability located in plasma membrane; 0.2000 probability located in endoplasmic reticulum (membrane); 0.1000 probability located in mitochondrial inner membrane; 0.0000 probability located in endoplasmic reticulum (lumen)
<b>SignalP analysis:</b>	Cleavage site between residues 49 and 50

A search of the NOV23a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 23D.

<b>Table 23D. Geneseq Results for NOV23a</b>				
<b>Geneseq Identifier</b>	<b>Protein/Organism/Length [Patent #, Date]</b>	<b>NOV23a Residues/ Match Residues</b>	<b>Identities/ Similarities for the Matched Region</b>	<b>Expect Value</b>
AAV55809	Human RCE1 (farnesyl-directed endopeptidase) sequence - Homo sapiens, 329 aa. [WO9961628-A2, 02-DEC-1999]	1..96 1..96	95/96 (98%) 96/96 (99%)	5e-51
AAW89181	Human RCE1 (hRCE1) polypeptide - Homo sapiens, 329 aa. [EP887415-A2, 30-DEC-1998]	1..96 1..96	95/96 (98%) 96/96 (99%)	5e-51
AAW98105	Guman ras carboxy-terminal processing protein - Homo sapiens, 338 aa. [WO9914343-A1, 25-MAR-1999]	1..96 10..105	95/96 (98%) 96/96 (99%)	5e-51
AAV26897	Human farnesylated--protein converting enzyme 2 protein - Homo sapiens, 329 aa. [WO9935275-A1, 15-JUL-1999]	1..96 1..96	95/96 (98%) 96/96 (99%)	5e-51
AAU03600	Human ras converting endoprotease (RCE) - Homo sapiens, 329 aa. [US6261793-B1, 17-JUL-2001]	1..96 1..96	94/96 (97%) 96/96 (99%)	1e-50

5 In a BLAST search of public sequence databases, the NOV23a protein was found to have homology to the proteins shown in the BLASTP data in Table 23E.

<b>Table 23E. Public BLASTP Results for NOV23a</b>				
<b>Protein Accession</b>	<b>Protein/Organism/Length</b>	<b>NOV23a Residues/</b>	<b>Identities/ Similarities</b>	<b>Expect Value</b>

Number		Match Residues	for the Matched Portion	
Q9Y256	CAAX prenyl protease 2 (EC 3.4.22.-) (Prenyl protein-specific endoprotease 2) (Farnesylated-proteins converting enzyme 2) (FACE-2) (hRCE1) - Homo sapiens (Human), 329 aa.	1..96 1..96	95/96 (98%) 96/96 (99%)	1e-50
P57791	CAAX prenyl protease 2 (EC 3.4.22.-) (Prenyl protein-specific endoprotease 2) (Farnesylated-proteins converting enzyme 2) (FACE-2) - Mus musculus (Mouse), 329 aa.	1..96 1..96	89/96 (92%) 90/96 (93%)	8e-46
Q9CSF8	Ras and a-factor-converting enzyme 1 homolog ( <i>S. cerevisiae</i> ) - Mus musculus (Mouse), 314 aa (fragment).	28..96 13..81	63/69 (91%) 65/69 (93%)	2e-31
Q8SZZ3	LD46418p - <i>Drosophila melanogaster</i> (Fruit fly), 302 aa.	38..86 30..78	24/49 (48%) 31/49 (62%)	2e-06
Q9U1H8	CAAX prenyl protease 2 (EC 3.4.22.-) (Prenyl protein-specific endoprotease 2) (Farnesylated-proteins converting enzyme 2) (FACE-2) (Severas protein) - <i>Drosophila melanogaster</i> (Fruit fly), 290 aa.	38..86 18..66	24/49 (48%) 31/49 (62%)	2e-06

PFam analysis predicts that the NOV23a protein contains the domains shown in the Table 23F.

Table 23F. Domain Analysis of NOV23a			
Pfam Domain	NOV23a Match Region	Identities/ Similarities for the Matched Region	Expect Value
No Significant Matches Found			

#### Example 24.

5 The NOV24 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 24A.

Table 24A. NOV24 Sequence Analysis			
	SEQ ID NO: 299	1701 bp	
NOV24a, CG153033-01 DNA Sequence	ATGGCTGGACCAGGCAAAGAGGGTGTGGTGTGGTGGGAAGAAAAATCGATGGGACAAC TGAGGGAAGAAGATAACATTGAGCTGAATGAAGAAGGAAGGCCGGTGCAGACGTCACG GCCAAGCCCCCACTCTGCGACTGCCACTGCTGCGGCCTCCCCAAGCGTTACATCATT GCTATCATGAGTGGGCTGGGATTCTGCATTTCCTTTGGGATCCGGTGCAATCTTGGAG		

	TTGCCATTGTGGAATGGTCAACAATAGCACCGTATATGTTGATGGAAAACAGACAGC ACAGTTTAACTGGGATCCAGAAACAGTGGGCCTTATCCATGGATCTTTTTTCTGGGGC TATATTATGACACAAATTCAGGTGGTTTCATTTCAAACAAGTTTGCTGCTAACAGGG TCTTTGGAGCTGCCATCTTCTTAACATCGACTCTGAACATGTTTATTCCCTCTGCAGC CAGAGTGCATTACGGATGCGTCATGTGTGTGAGAATTCTGCAAGGTTTAGTGGGTGTG ACCTACCCAGCCTGCCATGGGATGTGGAGTAAGTGGGCACCACCTTTGGAGAGAAGCC GACTGGCCACAACCTCTTTTTGTGGTTCCATGCAGGGGCAGTGGTTGCCATGCCCTT GGCTGGGGTGTGGTGCAGTACATTGGATGGTCCTCTGTCTTTTATATTTATGGTATG TTTGGGATTATTTGGTACATGTTTTGGCTGTTGTCAGGCCTATGAGTGCCCAGCAGCTC ATCCAACAATATCCAATGAGGAGAAGACCTATATAGAGACAAGCATAGGAGAGGGGGC CAACGTGGTGTAGTCTAAGTGTAATAATTTAGTACCCCATGGAAGAGATTTTTCATATCT TTGCCGGTTTATGCAATCATTGTGGCAAATTTTTGCAGAAGCTGGACCTTTTATTTC TCCTCATAAGTCAGCCTGCTTATTTTGAAGAGGTCTTTGGATTTGCAATAAGTAAGGT AGGTCTCTTGTGTCAGCAGTCCCACACATGTTTATGACAATCGTTGTACCTATTGGAGGA CAATTGGCTGATTATTTAAGAAGCAGACAAATTTTAACCAACAACCTGCTGTGAGAAAAA TCATGAACCTGTGGAGGTTTTGGCATGGAGGCAACCTTACTCCTGGTGGTTGGCTTTTC GCATACCAAAGGGGTGGCTATCTCCTTTCTGGTACTTGCTGTAGGATTTAGTGGCTTC GCTATTTTCAGGTTTTAATGTCAACCACCTGGACATTGCCCCACGCTATGCCAGCATTTC TCATGGGGATCTCAAACGGAGTGGGAACCTCTCTGGAATGGTCTGTCCCCTCATTGT CGGTGCAATGACCAGGCACAAGACCCGTGAAGAATGGCAGAAATGTGTTCTCATAGCT GCCCTGGTGCATTACAGTGGTGTGATCTTCTATGGGGTCTTTGCTTCTGGGGAGAAAC AGGAGTGGGCTGACCCAGAGAATCTCTCTGAGGAGAAATGTGGAATCATTGACCAAGC CGAATTAGCTGAGGAGATAGAACTCAACCATGAGAGTTTTGCGAGTCCCAAAAAGAAG ATGTCTTATGGAGCCACCTCCCAGAATTGTGAAGTCCAGAAGAAGGAATGGAAGGAC AGAGAGGAGCGACCTTGATGAGGAAGAGCTGACATCCTACCAGAATGAAGAGAGAAA CTTCTCAACTATATCCTAA		
	ORF Start: ATG at 1		ORF Stop: TAA at 1699
	SEQ ID NO: 300	566 aa	MW. at 62488.6kD
NOV24a, CG153033-01 Protein Sequence	MAGPGKEGVVWEEKSMGQLREEDNIELNEEGRPVQTSRPSPLCDCHCCGLPKRYII AIMSGLGFCISFGIRCNLGVAIIVEMVNNSTVYVDGKQTAQFNWDPETVGLIHGSFFWG YIMTQIPGGFISNKFAANRVFGAAIFLTSTLNMFI PSAARVHYGCVMCVRILQGLVGV TYPACHGMWSKWAPPLERSRLATTSFCGSYAGAVVAMPLAGVLVQYIGWSSVFYIYGM FGIIWYMFLLQAYECPAAHPTISNEEKTYIETSIGEGANVVSLSVKFSTPWKRFTS LPVYAIIVANFCRSWTFYLLLLISQPAYFEEVFGFAISKVGLLSAVPHMVMITVVP IGGQLADYLRSRQILTTTAVRKIMNCGGFGMEATLLLVVGFSTKGVAFISFLVLAVGFSGF AISGFNVNHLDIAPRYASILMGISNGVGTLSGMVCPLIVGAMTRHKTREEWQNVFLIA ALVHYSGVIFYGVFASGEKQEWADPENLSEKCGIIDQDELAEEIELNHESFASPKKK MSYGATSQNCEVQKKEWKQRGATLDEEELTSYQNEERNFSTIS		

Further analysis of the NOV24a protein yielded the following properties shown in Table 24B.

Table 24B. Protein Sequence Properties NOV24a	
PSort analysis:	0.6000 probability located in plasma membrane; 0.4000 probability located in Golgi body; 0.3000 probability located in endoplasmic reticulum (membrane); 0.3000 probability located in microbody (peroxisome)
SignalP analysis:	No Known Signal Sequence Predicted

5 A search of the NOV24a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 24C.

Table 24C. Geneseq Results for NOV24a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV24a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAU99329	Human transporter protein - Homo sapiens, 589 aa. [US2002082190-A1, 27-JUN-2002]	4..566 16..589	553/575 (96%) 555/575 (96%)	0.0
ABB07689	Rat glutamate transporter VGLUT3 amino acid sequence - Rattus sp, 860 aa. [WO200208384-A2, 31-JAN-2002]	4..566 24..601	509/580 (87%) 532/580 (90%)	0.0
AAM79273	Human protein SEQ ID NO 1935 - Homo sapiens, 582 aa. [WO200157190-A2, 09-AUG-2001]	4..530 11..549	413/542 (76%) 473/542 (87%)	0.0
AAO13870	Human polypeptide SEQ ID NO 27762 - Homo sapiens, 567 aa. [WO200164835-A2, 07-SEP-2001]	24..528 38..551	404/514 (78%) 450/514 (86%)	0.0
AAW70500	Human sodium-lithium countertransporter BNPI - Homo sapiens, 560 aa. [WO9838203-A1, 03-SEP-1998]	24..528 31..544	403/514 (78%) 449/514 (86%)	0.0

In a BLAST search of public sequence databases, the NOV24a protein was found to have homology to the proteins shown in the BLASTP data in Table 24D.

Table 24D. Public BLASTP Results for NOV24a				
Protein Accession Number	Protein/Organism/Length	NOV24a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
CAD30553	Vesicular glutamate transporter 3 - Homo sapiens (Human), 589 aa.	4..566 16..589	553/575 (96%) 555/575 (96%)	0.0
CAD37138	Vesicular glutamate transporter 3 - Rattus norvegicus (Rat), 588 aa.	4..566 16..588	510/575 (88%) 533/575 (92%)	0.0
Q9JI12	Differentiation-associated Na-dependent inorganic phosphate cotransporter - Rattus norvegicus (Rat), 582 aa.	4..561 11..579	421/573 (73%) 487/573 (84%)	0.0
Q920B7	Vesicular glutamate transporter 2 - Mus musculus (Mouse), 582 aa.	4..530 11..549	417/542 (76%) 475/542 (86%)	0.0

CAD52142	SI:PACKT73.2 (novel protein similar to solute carrier family 17 (sodium-dependent inorganic phosphate cotransporter), member 6 (SLC17A6)) - Brachydanio rerio (Zebrafish) (Danio rerio), 584 aa.	2..530 8..550	418/545 (76%) 472/545 (85%)	0.0
----------	--	------------------	--------------------------------	-----

PFam analysis predicts that the NOV24a protein contains the domains shown in the Table 24E.

Table 24E. Domain Analysis of NOV24a			
Pfam Domain	NOV24a Match Region	Identities/ Similarities for the Matched Region	Expect Value
sugar_tr	64..488	72/506 (14%) 262/506 (52%)	0.04

### Example 25.

The NOV25 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 25A.

5

Table 25A. NOV25 Sequence Analysis			
	SEQ ID NO: 301	3374 bp	
NOV25a, G153818-01 DNA Sequence	GCAATCATGAAGGACAGCGGGGACTCCAAGGACCAGCAACTCATGGTGGCGCTTCGGG TCCGGCCCATCAGCGTGGCAGAGCTGGAGGAAGGAGCTACCTCATCGCCATAAAGT GGATGAGCAGCATTTACCTGCTGCCACCCCCCTCTGCTCCCGGGGTGCTGTAGAGCCA GGCTCAAAGCTGCAAAGGGCCACTGGAGCAGTTCCTCACAGCCCTCTCAGCTGCGAG TGGAGATCCCCAAGCCAGCGTGTGACCTCATCCCTCAGCCAGCTGCCTGTGGCTCT TTGCTCTGTCCCAGGCTCTGCCCTGGAGGGGGCCCGGGTTCCAGGTGACCGTGGGC CTCCCTCTGGGGACCTTGCAGATGGTGGTTCTCATGGACCAATGGAGGATCCCGACG ACATCTGCGGGCGCATCGTCCCGGAGAAAGTCTACCTGTTCGACGTGGCCTTTGA CTTCACCGCCACCCAGGAGATGGTGTATCAGGCCACCACCAAGAGCCTCATCGAGGGC GTCATCTCAGGCTACAATGCCACTGTCTTTGCCATATGGCCACAGGTAAGGGGAATGC CAGACTTGTGCGAGACAGCAATGATCTGCTGTGGGAAAACCTACACCATGCTGGGCAC AGACCAGGAGCTGGCATCTATGTTTACAGCCCTCAACGACCTCTTCCGTGCCATCGAG GAGACCAGCAATGACATGGAGTATGAGGTCTCCATGTCTTACCTGGAGATCTACAATG AGATGATCCGGGACCTGCTGAACCCCTCCCTGGGCTACCTGGAGCTGCGGGAGGACTC TAAGGGGGTGATCCAGGTGGCCGGCATCACCGAAGTCTCCACCATCAATGCCAAGGAG ATCATGCAGCTGCTGATGAAGGGGAACCGGCAGAGGACCCAGAGCCACGCGCCGCA ACCAGACGTCTCCCGCTCCACGCGGTACTGCAGGTGACCGTGCGCCAGCGCAGCCG GGTCAAGAACATCTTGCAGGAGGCGCAGGGCCGCTGTTCATGATCGACCTGGCTGGC TCAGAGCGCGCTCGCAGACACAGAATCGTGGGCAGCGTATGAAGGAGGGGGCCACA TCAACCGCTCACTGCTGGCACTGGGCACTGCATCAACGCCCTGAGCGACAAGGGTAG CAACAAGTACATCAACTATCCGCAGACAGCAAGCTACCCGGCTCTGAAGGACTCTCTG GGAGGAAACAGCCGACAGTGATGATCGCTCACATCAGTCTCGCAGCAGTGCCCTTCG AGGAGTCCCGGAACACCTGACCTACGCCGGCCGGGCCAAGAACATTAAGACTAGGGT GAAGCAGAACCTCCTGAACCTCTCCTACCACATCGCCAGTACACAGCATCATCGCT GACCTGCGGGGCGAGATCCAGCGACTCAAGCGCAAGATTGATGAGCAGACTGGGCGGG GCCAGGCCCGGGCCGGCAGGATCGGGGTGACATCCGCCACATCCAAGCTGAGGTCCA GCTGCACAGCGGGCAGGGTGAGAAGGCTGGCATGGGACAGCTTCGGGAGCAGCTCGCC AGCGCCTTCCAGGAGCAGATGGATGTGCGGAGGCGCCTGCTGGAGCTGGAGAACCGCG		

	<p>CCATGGAGGTCCAGATTGACACCTCCCGACACCTGCTCACCATCGCCGGCTGGAAGCA  TGAGAAGTCCCGCGGGCCCTCAAATGGCGGGAGGAGCAGCGAAAGGAGTGCTACGCT  AAGGACGACAGCGAGAAGGACTCAGACACAGGTGATGACCAACCAGACATCCTGGAGC  CACCCGAGGTGGCCGAGCCCGGGAGAGCATTGCAGCCCTGGTGGACGAGCAGAAGCA  ACTGCGCAAGCAGAAGGTGTCCAGGGTTTGGGGGGACAAGGAGAGTGGGTTAGGGGA  CAGGATGCTGACCTGCGCCTCCTGCAGCTGGCGCTGGAGCAGCGCTGCCGGGAGCTGC  GCGCGCGGGGCGCGCCTGGAGGAGACGCTGCCGCGGCGCATCGGCTCCGAGGAGCA  GCGCGAGGTGCTCAGCCTGCTGTGCCGCTGCACGAGCTCGAGGTGAGAACACCGAG  ATGCAGTCGCACGCGCTGCTCCGCGACGCTGCGCTCCGCCACCGCCACGAGGCCGTGC  GCCGCTGGAGCAGCACCGCAGTCTCTGCGACGAGATTATCCAGGGCCAGCGGCAGAT  CATCGACGCACTACAACCTGGCCGCTCCCGCAGCGCCTGGAAGAGCTCTACGAACTG  TACCTGCGGGAGCTGGAGGAGGGCAGCCTGGAGCAGGCCACCATCATGGACCAAGTGG  CCTCCAGGGCCCTGCAGGACAGCTCCTTGCCCAAAATTACCCAGCAGGAACCTCACT  GACCCAGATTCTGACCTGGAGAGTGTGAAGACATTGAGCTCTGATGCCAGCAGCTG  CAGAACAGCGCCCTCCCTCCCTCAGCACAGAGAGTGAAGGCCACCACGTGTTCAAGG  CTGGTACTGGGGCCTGGCAGGCAAAAAGCTCCTCTGTGCCCAACCCACCTCCCATCCA  GCTCGGCAGCCTGGTGACGACGAGGAGGCGCCGCTCAGGACAGCCTGGGCAGCTGGATC  AACTCTTCCCTGACAGCAGTGAAGACCTGTGCGAGATCCCTTGTCCACAAAGAGA  GGAAGGAGATCTGACTGGCACCAGTGCATCTGGTGGAAGGCCGCCGCGCGCGCTC  GCGGGCCCTGGGAACCGAGGGGCGACCTGCTGGCACCCGCGACAGAGCGCAGCAGC  CTGTCCCTGCACTCACTGAGCGAGGGCGACGATGCGCGGCCACAGGCCCACCTGGCCT  GCAAGCGGCCGCCAGCCCCACACTACAGCATGTGCCAGTGAGGACAACCTGTCCAG  CAGCACGGGCGAGGCCCCGTCCCGGCGAGTGGACATCATGGGGACGGCCCCAGGCCC  TGGCTGCGTGCCAGAAGAAAAGCCTGGGCAAGAAAAGGGAGGAGTGCCTGGAGGCAA  AGAGAAGGAAGCGAGGTCCCGATCCTTCGAGGTACCGGGCAAGGGCTCTCCACCC  CAAGACACACCTCCTGGGGCCCCATCAGGCGGAGCGCATCTCGGACCACAGGATGCCA  GTGTGCAGGCAACCCAGCCCTGGTATCCGGCATCTGGGAAAGGTACGCTACCTTTGG  CCAAAGTCAAACCTCCCTCAAGCCAGAACACGGGCCCGGGGACTCCTCACCCCTGGC  TGTTCCTCCCAACCCAGGTGGTGGTTCGACGGGTACCCGTGGGCCCCGCTGCCC  CATGGCACAAAGCAGCCATGGCAAAGATGGATGCTCCCGGCATAACTGAGGGGGCCTGC  CTGGAAGTGG</p>
	<p>ORF Start: ATG at 7</p>
	<p>ORF Stop: TGA at 3352</p>
	<p>SEQ ID NO: 302</p>
	<p>1115 aa</p>
	<p>MW at 123442.0kD</p>
<p>NOV25a, CG153818-01 Protein Sequence</p>	<p>MKDSGDSKQQLMVALRVRI SVAELEEGATLIAHKVDEQHLPAATPLCSRGAPEPGS  KLQRATGAVPSQPSQLRVEI PKPSVLTSSLTQLPVALCSVPGSALEGARSGVTVGLP  LGTLMVVLMDPMDPDILRAHRSREKSYLFDVAFDFTATQEMVYQATTKSLIEGVI  SGYNATVFAYGPQVRGMPDLCEAMICCGKTYTMLGTDQEPGIYVQTLNDLFRAIET  SNDMEYEVMSYLEIYNEMIRDLLNPSLGYLELREDSKGVIQVAGITEVSTINAKIEM  QLLMKGNRQRTQEPTAANQTSRSHAVLQVTVRQSRVKNILQEAQGRFLMIDLAGSE  RASQTQNRGQRMKEGAHINRSLALGNCINALSDKGSNKYINVRDSKLTRLKDSLGG  NSRTVMIAHISPASSAFEESSNTLTLYAGRAKNIKTRVKQNLNVSYHIAQYTSIIADL  RGEIQRLLKRIDEQTGRGQARGRQDRGDIRHIAEVQLHSGQGEKAGMGQLEQLASA  FQEQMDVRRRLLELENRAMEVQIDTSRHLLTIAGWKHEKSRRALKWREEQRKECYAKD  DSEKDSDTGDDQPDILEPPEVAAARESI AALVDEQKQLRKQKVS RVWGDKESGFRGQD  ADLRLQLALEQRCRELRARGRRLEETLPRRIGSEEQREVL SLLCRVHELEVENTEMQ  SHALLRDGALRRHRHEAVRLEQHRSLCDEIIQGRQIIDADYNLAVPQRL EELYEVYL  RELEEGSLEQATIMDQVASRALQDSSLPKITPACTSLTPDSDL SVKTLSSDAQHLQN  SALPPLSTESEGHVFKAGTGAWQAKSSSVTPPPIQLGSLVTQEAPAQDSLGSWINS  SPDSSENLSEIPLSHKERKEILTGTCKIWKAAARRRSRALGTGRHLLAPATERSLSLS  LHSLSEGDDARPPGPLACKRPPSPTLQHAASEDNLSSTGEAPSRAVGHG DGRPWL  RGQKKS LGKKREESLEAKRRKRRSRSEFVTGQGLSHPKTHLLGPHQAERISDHRMPVC  RHPAPGIRHLGKVTLPLAKVKLPSPQNTGPGDSSPLAVPPNPGGSSRRATRGPRLPHG  TSTHGKDGCSRHN</p>

Further analysis of the NOV25a protein yielded the following properties shown in Table 25B.

<b>Table 25B. Protein Sequence Properties NOV25a</b>	
<b>PSort analysis:</b>	0.9800 probability located in nucleus; 0.3000 probability located in microbody (peroxisome); 0.1000 probability located in mitochondrial matrix space; 0.1000 probability located in lysosome (lumen)
<b>SignalP analysis:</b>	No Known Signal Sequence Predicted

A search of the NOV25a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 25C.

<b>Table 25C. Geneseq Results for NOV25a</b>				
<b>Geneseq Identifier</b>	<b>Protein/Organism/Length [Patent #, Date]</b>	<b>NOV25a Residues/ Match Residues</b>	<b>Identities/ Similarities for the Matched Region</b>	<b>Expect Value</b>
AAO21658	Protein fragment of the motor domain HsKip3b - Homo sapiens, 299 aa. [US6368841-B1, 09-APR-2002]	111..442 1..299	289/332 (87%) 289/332 (87%)	e-155
AAM50137	Human kinesin motor protein HsKip3b motor domain - Homo sapiens, 299 aa. [US6294371-B1, 25-SEP-2001]	111..442 1..299	289/332 (87%) 289/332 (87%)	e-155
ABB64748	Drosophila melanogaster polypeptide SEQ ID NO 21036 - Drosophila melanogaster, 728 aa. [WO200171042-A2, 27-SEP-2001]	140..816 68..684	259/692 (37%) 379/692 (54%)	e-106
ABB07410	Human kinesin motor protein, HsKip3A - Homo sapiens, 864 aa. [WO200196593-A2, 20-DEC-2001]	140..483 64..395	161/346 (46%) 229/346 (65%)	3e-81
AAU76957	Novel human kinesin motor protein, HsKip3d - Homo sapiens, 898 aa. [WO200212268-A1, 14-FEB-2002]	140..537 68..444	171/400 (42%) 254/400 (62%)	3e-79

5 In a BLAST search of public sequence databases, the NOV25a protein was found to have homology to the proteins shown in the BLASTP data in Table 25D.

<b>Table 25D. Public BLASTP Results for NOV25a</b>				
<b>Protein Accession Number</b>	<b>Protein/Organism/Length</b>	<b>NOV25a Residues/ Match Residues</b>	<b>Identities/ Similarities for the Matched Portion</b>	<b>Expect Value</b>

BAC04386	CDNA FLJ37300 fis, clone BRAMY2015782, moderately similar to KINESIN-LIKE PROTEIN - Homo sapiens (Human), 548 aa.	90..637 11..544	510/549 (92%) 512/549 (92%)	0.0
Q9VFN0	CG9913 protein - Drosophila melanogaster (Fruit fly), 728 aa.	140..816 68..684	259/692 (37%) 379/692 (54%)	e-105
CAD49067	Kinesin, putative - Plasmodium falciparum, 1669 aa.	121..478 955..1304	191/363 (52%) 252/363 (68%)	4e-95
O14343	Kinesin-like protein 5 - Schizosaccharomyces pombe (Fission yeast), 883 aa.	7..486 2..437	195/485 (40%) 276/485 (56%)	1e-83
Q9SCJ4	Kinesin-like protein - Arabidopsis thaliana (Mouse-ear cress), 813 aa.	89..716 13..548	217/631 (34%) 338/631 (53%)	4e-83

PFam analysis predicts that the NOV25a protein contains the domains shown in the Table 25E.

Table 25E. Domain Analysis of NOV25a			
Pfam Domain	NOV25a Match Region	Identities/ Similarities for the Matched Region	Expect Value
kinesin	140..186	22/54 (41%) 38/54 (70%)	2.1e-10
kinesin	203..468	126/319 (39%) 212/319 (66%)	2.3e-89

### Example 26.

5 The NOV26 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 26A.

Table 26A. NOV26 Sequence Analysis			
	SEQ ID NO: 303	13734 bp	
NOV26a, CG154435-01. DNA Sequence	GTTCTTGCACTCTTAACAGATACACAGTGTAAAGAAAGGCCAAGATGACAATGGCCC CGGACGTCAGACTAGAGTATCTGGAGGAAGTTGCCTCCATCGTCCTGAAGTTCAGCC GGACAAGTGGAGCAAGCTGATAGGCGCCGAGGAGAACGTGGCCCTGTTACAGAGTTTC TTTGAAAAGCCCGACGTCCAGGTGCTGGTGCTGACGCTCAATGCAGCCGGCATGATCA TACCCTGCCTGGGCTTCCCCAGTCCCTCAAGTCCAAAGGGGTTTACTTCATCAAGAC AAAGTCCGAGAACATCAACAAGGACAACACAGGGCCCGGCTCCTTTACGGCGACATC AGCCCCACACCCGTGGACCAGCTGATCGCGGTGGTGGAGGAGTCTCTTCTCTGT TAAACCAAAGTGAGAACATGGCTGGATGGCCCCAGGTGGTCTCGGAAGACATCGTGAA GCAGGTCCACAGGCTGAAGAATGAAATGTTGTGATGAGTGGCAAGATCAAAGGCAAA ACCTTGCTGCCTATTCCGGAGCACCTGGGCAGCCTGGATGGCACGCTGGAGTCCATGG AGAGGATCCCCCTTCACTGGACAACCTTGCTCCTGCAGCCATTGAAACCACCATCAT CGACTGGTCCCACCAGATCCGGGATGTGCTGAGCAAAGACTCAGCCCAGGCGTGCTG		

GATGGGCTGCACCCCTGCCCAAGTGGAGTTCGAGTTCTGGGACACTCGGCTGCTGA  
 ACCTCAAGTGCATCCATGAACAGCTAAACAGACCCAAAGTGAACAAGATTGTTGAGAT  
 CCTAGAGAAAGCCAAAGCTGCTACTGGCCAGCCCTGCAAAACGTTTACACCAACGTC  
 ACTGAAGGGCTGAAGGAAGCCAACGACATCGTGCTCTATTGAAGCCCTACGGATCC  
 TGCTGGAGGAGATGGAACAAGCCGACTTCACGATGCTCCCCACCTTCATTGCCAAGGT  
 GCTGGACACCATCTGCTTCATCTGGGCCACCTCTGAGTACTATAACACACCTGCCAGG  
 ATCATCGTCATCCTGCAGGAGTTCTGCAACCAAATCATCGAGATGACACGAACCTTCC  
 TGAGCCCGGAAGAGGTGCTGAAGGGCTGCAAGGTGAAATCGAGGAAGTCTTGAGTGG  
 CATCTCCCTGGCTGTAAATGTGCTGAAGGAGCTTACCAGACGTACGACTTCTGCTGC  
 GTGAACATGAAGCTTTTCTTTAAGGACAAAGAGCCCGTGCCTTGGGAATTCCCTTCTT  
 CTCTTGCTTTTCCAGGATAAAATTCCTTCTTCCAGCGCATCCAGACCATTGAGGAACCT  
 CTATAAAACAGCAATTGAGTTTCTGAAGCTGGAGAAAATCGAGCTTGGGGGCGTGCGT  
 GGGAACCTCCTCGGGAGCCTGGTGACCCGTATCTATGATGAGGTCTTTGAGCTGGTGA  
 AGGTTTTTGCCGACTGCAAAATATGATCCCTTGACCCCTGGAGACTCGAATTTTGACCG  
 TGATTATGCTGATTTGAGATCAAAATCCAAGACCTGGATAGGAGGCTGGCCACGATC  
 TTTTGCCAAGGATTGATGACTGCAGCTGTATCAAGTCTCCGCAAAGCTCCTGTACA  
 TGTGTGGGGGCTCATGGAGCGGCCCTGATTCTTGCCGAGGTGGCGCCAGGTATTC  
 AGTCATGCTGGAGCTGTTGACGCTGAGCTAGACAATGCTAAGATCTTGTACGATGCC  
 CAGATGGCGGCCCTCCGAGGAGGGGAACATCCCCTGATCCACAAAACATGCCTCCCG  
 TGGCCGGGCGAGCTCAAATGGAGCCTGGAGCTGCAGGAGAGGCTAGAGGTGTCCATGAA  
 ACACCTGAAGCACGTGCAACACCCGGTCATGCTGGAGCAGAGGCCAAGCTGACCTAT  
 CAGAAGTATGACGAGATGATGGAGCTGCTGAGGTGCCACCCGAGAGAATCTACCAGC  
 AGTGGGTGGCGGGCGTGGAACAGGACTGCCACTTTAACCTGGGGCAGCCGCTGATTCT  
 GCGGGACGCGCTAGCAACCTCATCCACGTCAACTTCAGCAAAGCGTTGGTGGCAGTT  
 CTGAGAGAAGTCAAGTATTTGAATTTCCAGCAACAGAAAGAGATTCCAGACAGTGCCG  
 AGAGTCTGTTCTCAGAGAACGAAACTTTCCGGAAGTTTGTGGGAACCTGGAGCTCAT  
 CGTTGGCTGGTATAATGAGATAAAGACTATAGTGAAGGCAGTAGAATTTCTACTAATA  
 AAGTCAGAACTGGAAGCAATTGATGTCAAGTTATTGAGCGCTGAAACGACATTATTCT  
 GGAATGGCGAAGGTGTGTTTCAGTACATTCAAGAGGTGCGAGAAATTCTGCACAACCT  
 GCAGAACAGGATGCAAAAGGCAAAACAAAATATAGAAGGAATTTCCAGGCTATGAAG  
 GACTGGTCCGCCAACCCGCTGTTTGAAGAAAGGACAATAAGAAAGAGGCCCTGTTAG  
 ACTTGATGGAAGAATTGCCAACCTCAACAAGCGCTACGAGCAGCTCAGGGATGCTGG  
 AGTGAAGATCCAAGCCATGGAACACGAGAACTATTACGGGCAGACACACTGAGCCTG  
 CCCTGGAAGGATTATGTCTATCTACATTGACGACATGGTCTTAGATGAATTTGACCAGT  
 TCATTGCAAAATCTCTGAGTTTCTAATGGACAACATGGTTATAGATGAGAGTATCGC  
 TCCCTGTTTGAGATCCGCATGGAGCTGGACGAGGATGGGCTGACCTTCAACCCGACC  
 CTGAGGTGGGCTCAGATCGCGCTTCTGGCACTGATCGAGGGCTGGTCAACGACA  
 TCTACAACGTAGCCAGGCTCATCCCTCGGCTGGCCAAGGACAGGATGAACACAGAT  
 GGACCTGGAAGATAACACAGACCTCATAGAGATGAGGGAGGAGGTGTCCAGCCTGGTC  
 ATCAATGCCATGAAGGAGGCCGAGGAGTACCAGGATTCTTTGAGAGGTACTCCTACC  
 TCTGGACGGACAACCTGCAGGAGTTTATGAAGAATTTCTGATATATGGGTGTGCAGT  
 CACTGCGGAGGACTTGGACACCTGGACAGATGACACCATCCCCAAGACACCGCCACC  
 CTGGCTCAGTTCCAGGAGCAGATCGACTCTACGAGAAGCTGTATGAGGAGGTGTCCA  
 AGTGCAGAGAACACCAAGGTGTTCCACGGCTGGCTGCAGTGCAGTGCAGGCTCCGCCCCCTCAA  
 GCAGGCCCTGCTCAGCACAATCCGGCGCTGGGGCTTCATGTTCAAGCGGCACCTGAGC  
 AACCCAGTCAACCAACAGCCTGGCTGACCTGGAAGCCTTCATGAAAGTCGCCAGAATGG  
 GCTTGACCAAGCCCTCAAGGAGGGGACTATGATGGGCTTGTGGAGGTGATGGGGCA  
 CCTGATGAAAGTCAAGGAGAGGCAAGCAGCCACGACAACATGTTTGAGCCCTGAAG  
 CAAACCATCGAGCTGCTCAAGACCTACGGGGAGGAGATGCCAGAGGAGATCCACTTGA  
 AGCTGCAGGAGCTGCCGGAGCACTGGGCAAATACCAAGAACTGGCCATTAGGTGAA  
 GCTGACCGTGGCAACCTCCAGGCCAACGAGGTCAGCATCCTGCGCGGAAATGCCAG  
 CAATTCGAGCTCAAGCAACATGAGTTTCAAGGAGAGGTTTCAAGCGCGAGGCCCGTTCT  
 CCTTCAGCGACCCCAACCCCTACAAGTCCCTGAATAAGCAACAAAGAGCATCTCCGC  
 CATGGAAGGCATCATGGAGGCGCTGTCCAAGTCCGGGGGCTGTCTGAGGTCCCGTCTC  
 CCAGACTACAAGCAGCTCAAGGCCTGCCACCGGGAGGTCCGCCTACTGAAGGAGCTCT  
 GGGACATGGTTGTTGTGGTAAATACCAGCATCGAGGACTGGAAGACCACCAAGTGGA  
 AGATATCAACGTTGAGCAGATGGACATAGATTGTAAGAAGTTTGCCAAGGACATGAGG  
 TCTTTGGACAAGGAGATGAAAACCTGGGATGCCTTCGTGGGGCTCGACAACACCGTGA  
 AAAACGTGATCAGTCCCTGCGTGCCGTGAGCGAGCTGCAGAACCTGCCATTCCGGGA  
 ACGCCACTGGCAGAGCTCATGCAAGGCCACCCAGGTGAAATTTAAATGTGCAAGAG  
 ACGACCTGGCAGATTTACTGCAGCTGAACCTCCACAGTTACGAGGATGAGGTCCGCA

ACATCGTGGACAAGGCCGTGAAGGAGTCGGGCATGGAAAAGGTGCTGAAAGCCCTGGA  
CAGTACCTGGAGCATGATGGAATTCCAGCACGAGCCGCCACCCGGGACAGGCACCATG  
ATGCTCAAGTCCAGCGAGGTGCTGGTGGAGACGCTGGAGGACAACAGGTGCAGCTGC  
AGAACCTGATGATGTCCAAGTACCTGGCCCCACTTCCTGAAGGAGGTGACAAGCTGGCA  
GCAGAAGCTGTCCACGGCGGACTCCGTCACTCCATCTGGTTTGAGGTCCAGCGAACC  
TGGAGCCACCTGGAGAGCATCTTCATCGGCTCCGAAGACATCCGCACCCAGCTCCCGG  
GGGACTCCAGCGCTTTGACGACATCAACCAGGAATTCAGGCCTTGATGGAAGATGC  
AGTGA AACACCCACGTGGTGGAGCCACCAGCAAACCCGGCCTCTACAATAAAGT  
GAGGCCCTGAAGAAGAGCTTGGCCATCTGTGAAAAGGCTTTGGCAGAGTATTTAGAGA  
CGAAAAGACTGGCTTTCCCCGGTTCTATTTGTCTCCTCGGCTGACCTCCTGGACAT  
TCTCTCCAATGGCAATGACCCCGTGGAGGTGAGCCGCCACCTGTCCAACTCTTCGAT  
AGCCTGTGTAAACTGAAGTTCCGGCTCGATGCCAGTGACAAACCTCTCAAGGTGGGCC  
TGGGAATGTACAGCAAGGAGGACGAGTACATGGTTTTTGATCAGGAATGCGACCTCTC  
GGGGCAGGTGGAAGTGTGGCTGAATCGAGTGTGGACCGAATGTGCTTACCTCCGG  
CACGAAATCCAGAGGCCGTGGTGACCTACGAAGAGAAGCCGAGGGAGCAGTGGATCC  
TGGACTACCCAGCCAGGTGGCCCTGACTTGACCCAGATCTGGTGGACGACCGAGGT  
GGGCCCTGGCATTGTGCCAGGCTGGAGGAAGGCTATGAAAACGCTATCAGAGATTATAAC  
AAAAAGCAGATTAGCCAGCTGAACGTACTCATCAGCTGCTCATGGGGAACCTCAACG  
CTGGCGACAGGATGAAGATCATGACCATCTGCACCATCGATGTGCACGCACGGGACGT  
GGTGGCCAAAATGATCGTGGCCAAGGTGGAGAGTTCTCAGGCCTTCACTGGCAGGCC  
CAGCTCCGGCATCGCTGGGACGAAGAGAAGCGACACTGCTTTGCCAACATCTGCGATG  
CCCAAATCCAGTATTCCTATGAGTATCTGGGCAACACGCCCGGCTGGTCAACCC  
ACTCACTGACAGGTGCTATATCACCTGACCCAGTCCCTCCATCTCATCATGGGTGA  
GCCCCTGCCGGCCCCGCTGGGACCGGCAAGACTGAGACGACCAAGGACCTGGGCAGAG  
CCCTGGGCACCATGTCTACGTCTTCAACTGCTCCGAGCAGATGGACTACAAGTCCCTG  
TGGAATATCTACAAGGGCCTGGCCCAGACGGGAGCCTGGGGCTGCTTTGACGAGTTT  
AATCGCATCTCAGTGAAGTCTGTCTGTGATTGCCGTGACGGTAAATGTGTCCAGG  
ATGCAATTCGGGGCCAAGAAAAAGCATTCAATTTCTGGGAGAGATCATAGGCCTCAT  
TCCCACCGTCGGTATCTTCATCACCATGAACCTGGGTACGCCGGACGCGCGGAGCTG  
CCTGAGAACCTAAAAGCCTTATTCAGGCCCTGTGCCATGGTGTGCCCGACTTCGAAC  
TGATATGTGAGATCATGCTCATGGCCGAGGGCTTTCTGGAAGCCCGCCTTCTGGCCAG  
GAAGTTCATCACCTGTACACCTTGTGCAAGGAGCTGCTCTCGAAGCAGGATCATTAC  
GACTGGGGCTTGAGAGCCATCAAGTCTGTGCTGGTGGTGGCCGGCTCCCTGAAGAGGG  
GCGACCCAGCCGGGCAGAGGACCAGGTGCTCATGCGGGCGCTGAGAGACTTCAACAT  
CCCCAAGATTGTGACAGACGACCTGCCCGTATTCATGGGACTGATCGGGACCTCTTC  
CCGGCTCTGGACGTGCCTCGGAAACGGGACCTGAATTTGAAAAGATCATCAAGCAGA  
GCATCGTGGAGCTCAAGCTGCAGGCGGAGGACAGCTTCGTGCTGAAGGTGGTGCAGCT  
GGAGGAGCTGCTGCAGGTCCGCCACTCCGTGTTTCATCGTCGGGAATCGGGACCGGC  
AAATCTCAGGTCTCAATCCCTCAACAAGACCTATCAGAACCTGAAGAGGAAGCCGG  
TCGCCGTGGACCTGGACCCCAAGGCCGTACCTGCGACGAGCTCTTTGGCATCATCAA  
CCAGTGACACAGGAATGGAAGATGGCCTGTTCTCCACCATCATGCGAGACCTGGCC  
AACATCACCCATGACGGCCCCAAGTGGATCATCCTTGACGGAGACATAGACCCCATGT  
GGATCGAGTCTCTCAACACAGTCATGGATGACAACAAGGTCTCACCCTGGCCAGCAA  
CGAGCGGATCCCCCTGAACCGCACCATGAGGCTGGTGTTCGAAATCAGCCACCTGAGG  
ACGGCCACCCAGCCACCGTTTCCAGAGCCGGCATCCTCTACATCAACCCAGCCGACC  
TGGGATGGAACCCGGTGGTGAAGCTGGATCGAGAGGCGCAAGGTGAGTCCGAGAA  
GGCCAACTGATGATCCTCTTTGACAAGTACCTGCCACGTGCCCTGGACAAGTTGCGC  
TTTGGGTTCAAGAAGATCACGCCAGTGCCCGAGATCAGGTGATCCAAACGATTCTGT  
ACCTGCTGGAGTGCCTGCTCAGCGAGAAGACCGTGCCCCCGGACTCCCCAGGGAGCT  
GTACGAGCTGTACTTCGTGTTACCTGCTTCTGGGCCCTTCGGTGGCGCCATGTTCCAG  
GACCAGCTTGTGGATTATCGAGTGGAGTTCAGTAAATGGTGGATCAACGAATTCAGA  
CTATCAAGTTCCTCGCAGGGAACGATTTTGACTACTACATTGATCCTGACACAAA  
AAAGTTCCTGCCCTGGACAGATAAAGTGCCCTCCTTTGAGCTGGATCCCGATGTCCCA  
CTGCAGGCCTCTTTGGTCCACACACGAAACCATCCGCATCCGCTACTTTCATGGACC  
TGCTCATGGAGAAGTCTGGCCGGTGATGCTGGTGGGGAACGCGGACGCGGCAAGTC  
GGTGTGATGGGGGACAAGCTGGAAAGCCTGAACACGGACAACCTACCTGGTGCAGGCT  
GTGCCCTTCAACTTCTACACGACCTCAGCCATGCTGCAGGGGGTGTGGAGAAGCCGC  
TGGAGAAGAAATCGGGGAGGAACACGGGCCCGCAGGCACTAAGAAGCTCGTCTACTT  
CATCGACGACATGAACATGCCCCGAGGTGGACAAGTATGGGACGGTGGCCCCGCACACC  
CTCATCCGGCAGCACATGGACCACCGGCACTGGTATGACAGACATAAGGTGACGTTAA  
AAGATATCCATAATTGTCTAGTACGTGGCCTGCATGAACCCACTTCCGGATCCTTAC

CATCGACTCCAGGCTTCAGCGCCATTCTGCGTGTGCTGTGAGCTTCCCCGGCCAG  
 GAGGCCCTCACCACCATCTACAACACAATCCTGACGCAGCACCTGGCCTTCCGCTCGG  
 TCTCCATGGCTATCCAGAGGATAAGCAGCCAGCTGGTGGCCGCGGCCCTGGCTTTGCA  
 TCAGAAAATCACGGCAACATTTCTTCCCACGGCCATTAAGTTTCATTATGTCTTCAAC  
 CTCAGGGACCTCTCCAATATTTTCCAGGACTCTTATTTTCCACAGCAGAAGTTCTGA  
 AAACCCCACTGGACCTCGTCCGCCTTTGGCTACATGAGACTGAACGAGTGTATGGTGA  
 CAAAATGGTTGACGAAAAAGACCAGGAAACATTGCATAGAGTCACCATGGCCTCCACC  
 AAGAAGTCTTTTGATGATCTTGGTGATGAACCTTATTTGCAAGCCAAATATCTTCT  
 GCCACTTTGCTCAAGGGATTGGCGATCCCAAATATGTTCTGTAAACCGACATGGCTCC  
 TCTGAACAAGCTCCTCGTGGACGCTCTGGACAGCTACAATGAAGTTAATGCAGTCATG  
 AATTTGGTGCTGTTTGAGGACGCCGTGGCTCACATCTGCAGGATTAATCGCATCCTGG  
 AGTCTCCCCGGGGGAATGCCCTGCTGGTGGGGGTGGGCGGCAGTGGCAAAACAGAGCCT  
 CTCCCGCTGGCAGCGTACATCAGCGGGCTTGACGTGTTTTCAGATCACCTCAAGAAG  
 GGCTACGGGATCCCCGACCTCAAGATTGACCTCGCTGCTCAGTACATAAAGGCTGCCG  
 TGAAGAAGCTTCCCTCGGTGTTCTGATGACAGACTCCCAGGTGGCCGAGGAGCAGTT  
 TCTGGTGCTGATCAATGACCTGCTGGCCTCAGGAGAGATCCCTGGGCTGTTTATGGAG  
 GACGAGGTGGAGAACATCATCTCCTCCATGCGACCCCAAGTCAAGTCCCTTGGCATGA  
 ATGACACTCGGGAAACATGTTGGAAGTTCTTCATCGAAAAAGTGCAGACAGCTCAA  
 GGTGATCCTGTGTTTCTCCCCGTGGGCTCCGTGCTGCGGGTACGAGCCAGAAAGTTC  
 CCAGCTGTGGTCAACTGCACGGCCATCGACTGGTTCACGAGTGGCCGGAAGATGCCG  
 TGGTGTCGCTCAGCGCCCGCTTCTGGAGGAGACTGAGGGGATTCCGTGGGAAGTCAA  
 GGCCCTCCATCAGCTTCTTCATGTCTACGTGCACACCACCGTCAACGAGATGTCCAGG  
 GTATACCTGGCTACTGAGAGGCGCTACAACCTACACCACACCCAAAACCTTTCTGGAGC  
 AGATCAAACCTGTACAGAACCTGCTGGCCAAGAAGAGAACGGAACTTGTGCCAAAAT  
 CGAGAGGCTGGAGAACGGCCTGATGAAGCTGCAGAGCACGGCTTCCCAGGTGGATGAT  
 TTGAAGCCAAAGTTGGCGATTTCAGGAGGCTGAGCTCAAGCAGAAGATGAGAGCGCAG  
 ACCAACTGATCCAGGTGGTCCGCATCGAGGCCGAGAAGTTCAGCAAAAGAGAAGGCCAT  
 TGCTGACCAGGAAGAAGTCAAGGTGAGGTCATCAATAAGAACGTCAGTGAAGCAA  
 AAGGCTGTGAAACAGACCTGGCCAAAGCAGAACCAGGCTGCTGGCAGCCAGGAGG  
 CTCTGGACACTCTGAATAAGAACAACTGACAGAGCTGAAGTCTTTGGGTCCCCGCC  
 GGATGCTGTGGTCAACGTCAACGCCGCCGTCATGATTCTGACCGACCTGGGGGCAAG  
 ATCCCCAAGGACAAGAGCTGGAAGGCGGCCAAGATCATGATGGGCAAGGTGGACACCT  
 TCCTAGACTCCCTGAAGAAGTTCGACAAGGAGCACATCCCTGAGGCTGCCTGAAGGC  
 CTTCAAGCCCTACCAAGGCAACCCGACGTTTCGACCCGAGTTCATCCGCTCCAAGTCC  
 ACGGCCGCGCCGCGCCTGTGCTCCTGGTGCATCAACATCGTCCGCTTACGAGGTCT  
 ACTGCGACGTGGCGCCCAAGAGGCAAGGCACTGGAGGAGGCTAATGCAGAGCTGGCAGA  
 GGCACAAGAGAAGCTGTCCCGGATCAAAAAACAAGATTGCCGAAGTTAACGCCAAGT  
 AGCAACCTAACCTCAGCGTTTGAAGAACACAGCTGAGAAAATCAAGTGTGAGCAAG  
 AGGCCGATGCCACGAACAGGGTGATCTTACTGGCGAACAGGCTGGTGGGGGATTAGC  
 ATCGGAAAACATCCGCTGGGCTGAGTCTGTGGAGAATTCAGGAGCCAGGGGGTACG  
 CTGTGTGGGACGTCCTGCTCATCTCTGCCCTTCGTGTCTACGTGGGCTACTTCACCA  
 AGAAATACCGGAATGAGCTGATGGAGAAATCTGGATCCCTTACATACATAACTTAA  
 GGTCCCATCCCGATCAGCAATGGCCTGGATCCCTTGAGCCTGCTGACAGATGACGCG  
 GACGTGGCCACCTGGAACAACAGGGCCTCCCCAGCGACCGCATGTCCACCGAGAATG  
 CCACATCCTGGGCAACACCGAGCGGTGGCCGCTGATCGTGGACGCCAGCTCCAAGG  
 AATCAAGTGGATCAAAAACAAATACAGGAGTGAATGAAAGCCATCCGCTGGGACAG  
 AAGAGCTACCTGGATGTATCGAGCAGGCCATCTCGGAAGGGGACACCTTGCTCATTG  
 AGAACATCGGCGAAACCGTGGACCCCGTGTGGACCTCTACTGGGCAGGAACACGAT  
 TAAAAAGGGAAAGTACATTAGATCGGTGACAAGGAGGTGGAGTACCACCCCAAGTTC  
 CGCCTGATCCTACACACCAAGTACTTCAACCCACACTACAAGCCAGAGATGCAGGCTC  
 AGTGACCCCTCATCAACTTCTGGTACCAGGGATGAGTTCGAGGACCAACTCTTGGC  
 CGCTGTGGTGGCCAAAGAGCGCCAGATCTGGAACAGCTGAAGGCAAACTCACCAG  
 TCTCAAAACGAATTTAAGATTGTTCTGAAAGAGCTGGAAGATTGCTCCTGGCCCGT  
 TGTGCGCTGCGTTCGGGGAACCTTCTGGGAGACACGGCCTTGGTGGAGAAATCGGAGAC  
 CACCAAGCACACAGCCAGCGAGATCGAGGAGAAGGTGGTGGAGGCAAAAATCACAGAA  
 GTTAAATCAACGAAGCGAGAGAGAACTACCGCCCGGCTGCGGAGAGGGCATCTCTGC  
 TCTACTTCATACTGAACGATCTCAACAAAATCAACCCGCTTACCACTTCTCCCTCAA  
 GGCTTCAACGTGGTGTGAGAAAGCCATCCAGAGGACACCCCTGCCAACGAGGTG  
 AAGCAGCGGTGATCAACCTGACGACGAGATCACTACTCCGTCTACATGACACGG  
 CCGGGGACTCTTCGAGAGGACAAACTCATTTCTGCGCAAGTTACGTTTCAGGT  
 CCTGTCCATGAAGAAGGAGCTGAACCCAGTGGAGCTGGATTCTCTCTGCGGTTCCCT

	<p>TTTAAGGCCGGAGTGGTCTCACCAGTGGACTTCCTCCAGCATCAAGGCTGGGGCGGGA  TCAAGGCCCTCTCGGAGATGGATGAGTTCAAAAATCTGGACAGTGACATCGAAGGATC  TGCCAAGCGCTGGA AAAAGCTGGTGGAGTCCGAAGCCCCGAGAAGGAGATCTTCCCC  AAGGAGTGAAGAACAAGACGGCCCTGCAGAAGCTGTGCATGGTGGCTGCCTGCGGC  CAGATCGCATGACCTACGCTATCAAGAACTTCGTGGAGGAAAGATGGCGCAAGTT  CGTGAAGGCCGGAGTGTGAGTTTCTAAGTCTACGAGGAGAGCAGCCCTCCACG  TCAATCTTCTTCATCTCTCCCCGGGGTTGACCCCTTGAAGACGTGGAAGCCCTGG  GAAAAAACTAGGGTTTACCATAGACAATGGAAAACTCCATAATGTGTCCCTGGGGCA  GGGACAAGAGGTGGTGGCTGAGAACGCCCTGGACGTGGCTGCAGAGAAAGGACACTGG  GTCATTCTGCAGAATATCCACCTGGTGGCCCGTGGCTGGGAACACTGGACAAGAAGC  TGGAGTGTCTACGACACGGGCAGCCATGAGGACTACCGGGTGTTCATCAGCGCGGAGCC  TGCCCCCAGCCCCGAGACCCACATCATCCCCAGGGCATTCTGGAGAACGCCATCAAG  ATCACCACGAGCCCCCAGGGCATGCACGCCAATTGCACAAGGCCCTGGACCTGT  TCACCCAGGACACCCCTGGAGATGTGCACCAAGGAGATGGAGTTCAAGTGCATGCTCTT  CGCCCTGTGCTACTTCCACGCTGTGGTGGCAGAGAGGCGCAAGTTTCGGCGCCAGGGC  TGGAACCGGTCTACCCCTTCAACAACGGGGACCTCACCATCTCCATCAACGTGCTCT  ACAACCTACCTGGAGGCCAACCCCAAGGTGCCCTGGGACGATCTCCGCTACCTTTTTGG  TGAAATCATGTATGGCGGCCACATCACAGATGACTGGGACCGTGGCTGTGCAGGACC  TACCTGGCTGAATACATCCGGACGGAGATGCTGGAGGGAGACCTCTGTGTCGGCCCCG  GCTTTCAGATCCCCCAACCTGGACTACAAGGTTTACCACGAATACATCGATGAGAA  CCTGCCCCCTGAGAGTCCCTATCTGTATGGCCTGCACCCCAACGCAGAGATTGGCTTT  CTGACGGTCACTCAGAGAAGCTGTTCGCACTGTCTGGAAATGCAGCCAAAAGAGA  CGGACTCGGGGGCAGGCACGGGAGTGTCCCGCAGGAGAGAAGTGAAGGCCGTGTGGA  CGACATCTGGAGAAGATTCCGGAGACTTTCAACATGGCTGAGATCATGGCAAAGGCA  GCGGAAAAGACCCCCCTACGTGGTAGTCGCCCTTCAAGAATGTGAAAGAATGAACATCC  TGACCAACGAAATGCGCCGTTTCGCTCAAGGAGCTGAACCTGGGGCTGAAGGGAGAAT  GACCATCAGCACCGACGTGGAAGATCTGTCCACGGCTCTCTTATGACACCGTGCCCT  GATACGTGGGTGGCCCGGCCCTACCCCTCCATGATGGGCCTGGCGGCCCTGGTACGCAG  ACCTGTGCTCCGCATCAGGGAATCGAGGCCCTGGACGACAGACTTTGCCCTGCCAC  CACCCTGTGGCTGGCCCGGCTTCTTCAACCCCCAGTCTCTCCTACGGCCATCATGCAG  TCCATGGCCAGGAAGAACGAGTGGCCCCCTGGACAAGATGTGTCTGTCTGTGAGGTGA  CCAAGAAAAACCGAGAGGACATGACCGCTCCTCCGCGAGAGGGCTCCTACGTGTACGG  ACTCTTCATGGAAGGGGCTCGCTGGGACACCCAGACTGGAGTCATCGCTGAAGCGCGG  CTGAAAGAGCTGACCCGGCCATGCCTGTCTCTCATCAAGGCCATTCTGTGGACC  GCATGGAGACCAAGAACATCTATGAGTGTCCCGTGTACAAAACACGCATCCGCGGGCC  CACCTATGTCTGGACCTTAACTTGAAGACCAAGAGAAGGCAGCGAAGTGGATCCTG  GCAGCCGTGGCGCTGCTCCTACAGGTTTAGCTCGCTCCTGCCTCAGAGCCACACTCC  CTGGGGCTGGACCACAACCTCAGCCCTTACCTGTGCACCTGTGACTTATTCTTTACAG  GAACCTGGTGGTGGTTTTCTGTTCTTAAATAATCAGGTGCTTTGTAACCAAGCACAT  CGGAACCAGAGGGTGGAGGTTGGTGTGGAAGAGGTGGGGCAGATTAAAGCCAGTGGAG  CCACTCAGCTGTGCCCATCCATTCTGTGCCCTGATGGCCACTGTGAGGCCCTGGTTTCAGG  CTTTGGGGAAGGCCCAATTCACAGCCAGCCAGAGGCAAGCATTC</p>
	<p>ORF Start: at 61</p>
	<p>ORF Stop: TAG at 13426</p>
	<p>SEQ ID NO: 304</p>
	<p>4455 aa</p>
	<p>MW at 508571.2kD</p>
<p>NOV26a, CG154435-01 Protein Sequence</p>	<p>DVRLEYLEEVASIVLKFKPKDKWSKLIGAEENVALFTEFFEKPDVQVLVLTINAAGMI I  PCLGFPQSLKSKGVYFIKTKSENINKDNYRRLLYGDISPTPVDQLIAVVEVLSSLL  NQSEN MAGWPQVSEDI VKQVHRLKNEMFVMSGKIKGKTL LPIPEHLGSLDGTLESME  RIPSSLDNLLLHAIETTTIDWSHQIRDVLSKDSQAALLDLHLPLPQVEFEFWDTRLN  LKCIHEQLNRPKNKIVEILEKAKSCYWPALQNVYTNVTEGLKEANDIVLYLKPLRIL  LEEME QADFTMLPTFIKVLDTICFIWATSEYNTPARIIVILQEFNCQIIE MTRTFL  SPEEVLKGLQGEIEEVLSGISLAVNVLKELYQTYDFCCVNMKLF FFKDKFPVPWEPSS  LAFSRINSFFQRIQTIEELYKTAIEFLKLEKIELGGVRGNLLGSLVTRIYDEVFELVK  VFADCKYDPLDPGDSNFD RYADFEIKIQDLDRRLATIFCQGFDDCSCKSSAKLLYM  CGGLMERPLILA EVA PRYSVMLELFD AELDNAKILYDAQMAAEEGNIPLIHKNMPPV  AGQLKWSLELQERLEVSMKHLKHVEHPVMSGAEAKLTYQYDEMME LLRCHREKIYQ  WVAGVDQDCHFNLGQPLILRDAASNLIHVNFSKALVAVLREV KYLNFQQQKEIPDSAE  SLFSENETFRKFVGNLELIVGWYNEIKTIKAVEFLLIKSELEAIDVKLLSAETTLFW  NGEGVFQYIQEVREILHNLQNRMQAKQNI EGISQAMKDSANPLFERKDNKKEALLD  LDGRIANLNKRYAAVRDAGVKIQAMENAE LFRADTL SLPWKDYVIYIDDMVLDEFDQF</p>

IRKSLSFLMDNMVIDESIAPLFEIRMELEDEDGLTFNPTLEVSGDRGFLALIEGLVNDI  
 YNVARLI PRLAKDRMNYKMDLEDNTDLIEMREEVSSLVINAMKEABEYQDSFERYSYL  
 WTDNLQEFMKNFLIYGCAVTAEDLDTWTDDTIPKTPPTLAQFQEQIDSYEKLVEEVSK  
 CENTKVFHGWLQDCRPFKQALLSTIRRWGFMFKRHLSNHVTNSLADLEAFMKVARMG  
 LTKPLKEGDYDGLVEVMGHLMKVKERQAATDNMFEPKQTIELLLKTYGEMPEEIHKL  
 LQELPEHWANTKKLAIQVKLTVAPLQANEVSILRRKCQFELKQHEFRERFRREAPFS  
 FSDPNPYKSLNKQKQKSI SAMEGIMEALS KSGGLFEVPVPDYKQLKACHREVRLKELW  
 DMVVVNTSIEDWKTWKWDINVEQMDIDCKKFAKDMRSLDKEMKTWDAFVGLDNTVK  
 NVITSLRAVSELQNP AIRERHWQQLMQATQVKFKMSEETTLADLLQLNLHSEYDEVERN  
 IVDKAVKESGMEKVLKALDSTWSMMEFQHEPHPRGTMTMLKSSEVLVETLEDNQVQLQ  
 NLMMSKYLAHFLKEVTSWQQLLSTADSVISIWFVEVQRTWSHLESIFIGSEDIRTQLPG  
 DSQRFDDINQEFKALMEDAVKTPNVVEATSKPGLYNKLEALKSLAICEKALAEYLET  
 KRLAFPRFYFVSSADLLDILSNGNDPVEVSRHLSKLFDSLCKLKRDLASDKPLKVG  
 GMYSKEDYEMVFDQECDSLQGVVWLNRLDRMCSTLRHEIPEAVVYEEKPREQWIL  
 DYPAQVALTCTQIWWTTEVGLAFARLEEGYENAIRDYNKKQISQLNLVITLLMGNLNA  
 GDRMKIMTICTIDVHARDVAKMIVAKVESSQAFWQALRHRWDEEKRHCFCANICDA  
 QIQSYEYLGNTPRLVITPLTDRCYITLTQSLHLIMGGAPAGPAGTGKTETTKDLGRA  
 LGTMVYVFNCSQMDYKSCGNIYKGLAQTGAWGCFDEFNRISVEVLSVIAVQVKCVQD  
 AIRAKKAFNFLGELIGLIPTVGIFITMNPYAGRAELPENLKALFPCAMVVPDFEL  
 ICEIMLMAEGFLEARLLARKFITLYTLCKELLSKQDHYDWGLRAIKSVLVVAGSLKRG  
 DPSRAEDQVLMRALRDFNIPKIVTDDLVPFMGLIGDLFPALDVPKRDLNFEEKIKQS  
 IVELKLQAEDSFVLKVVQLELLQVRHSVFIVGNAGSGKSQVLKSLNKTYQNLKRKPV  
 AVDLDPKAVTCDLFGIINPVTREWKDGLFSTIMRDLANITHDGPKWIIDGDIDPMW  
 IESLNTVMDDNKVLTASNERIPLNRTMLVFEISHLRTATPATVSRAGILYINPADL  
 GWNPVVSSWIERRKVQSEKANMILFDKYLPTCLDKLRFGFKKITPVPEITVIQTILY  
 LLECLLTEKTVPPDSPRELYELYFVFTCFWAFGGAMFQDQLVDYRVEFSKWWINEFKT  
 IKFPSQGTIFDYIDPDTKKFLPWTDKVPSFELDPDVPQLASLVHTTETIRIRYFMDL  
 LMEKSWPVMVLGNAGTGKSVLMGDKLESNTDNYLVQAVPFNFYTTSAMLQGVLEKPL  
 EKKSGRNYGPPGTKLVYFIDDMNMPEVDKYGTVAHPTLIRQHMDHRHWYDRHKLTLK  
 DIHNCQYVACMNPTSGSFTIDSRLQRHFCVFAVSFPQGEALTTIYNTILTQHLLAFRSV  
 SMAIQRISSQLVAAALALHQKITATFLPTAIKFHYVFNLRDLNSIFQGLLFSTAELVK  
 TPLDLVRLWLHETERVYGDKNVDEKQDQETLHRVTMASTKKFFDDLDELFAKPNIFC  
 HFAQGI GPKYVPVTDMAPLNKLVDVLDVLSYNEVNAMNVLVLFEDAVAHICRINRILE  
 SPRGNALLVGVGSGKQSLRLAAYISGLDVFOITLKKGYGIPDLKIDLAAQYIKAAV  
 KNPVSVFLMTDSQVAEEQFLVLINDLLASGEIPGLFMEDEVENIISMRPQVKSLGMN  
 DTRETCWKFFIEKVRRLKVLICFSPVGSVLRVRARKFPVAVNCTAIDWFHEWPEDAL  
 VSVARFLEETEGIPWEVKASISFFMSYVHTTVNEMSRVYLATERRYNTTPKTFLEQ  
 IKLYQNLLAKKRTLVAKIERLENGLMKLQSTASQVDDLKAKLAIQEAELKQKNESAD  
 QLIQVVGIEAEKVSKEKAIADQEEVKVEVINKNVTEKQKACETDLAKAEPALLAAQEA  
 LDTLNKNNLTLSKSGSPDAVVNVTAAMILTAPGGKIPKDKSKWAAKIMMGKVDTF  
 LDSLKKFKDEHIPEACLKAFKPYQGNPTFDPEFIRSKSTAAAGLCSWCINIVRFYEVY  
 CDVAPKRQALEEANAELAEAEKLSRIKNKIAELNANLSNLTSAFEKATAEKIKCQOE  
 ADATNRVILLANRLVGGLASENIRWAESEVENFRSQGVTLCDVLLISAFVSVGYFTK  
 KYRNELEMEKFWIPYIHNKLVPI PITNGLDPLSLLTDDADVATWNNQGLPSDRMSTENA  
 TILGNTERWPLIVDAQLOGIKWIKNKRSELKAIRLGQKSYLDVIEQAISEGDTLLIE  
 NIGETVDPVLDPLLRNTIKKGKYIKIGDKEVEYHPKFRILILHTKYFNPHYKPEMQAQ  
 CTLINFLVTRDGLDQLLAADVAKERPDLEQLKANLTKSQNEFKIVLKELEDSLLARL  
 SAASGNFLGDTALVENLETTKHTASEIEEKVVEAKITEVKINEARENYRPAERASLL  
 YFILNDLNKINPVYQFSLKAFNVFEKAIQRTTPANEVKQORVINLTDEITYSVMYTA  
 RGLFERDKLI FLAQVTFQVLSMKKELNPVELDFLLRFPFKAGVSPVDFLQHQWGGI  
 KALSEMDEFKNLSDIEGSAKRWKKLVESEAPEKEIFPKWKNTALQKLCMVRCRLP  
 DRMTYAIKNFVEEKMGSKFVEGRSVEFSKSYEESPSTSIFFILSPGVDPLKDVEALG  
 KKLGFITIDNGKLHNVS LGQGEVVAENALDVAEKGHWILQNIHLVARWGLTDLKKL  
 ECYSTGSHEDYRVFISAEPAPEPETHIIPQGIENAIKITNEPPTGMHANLHKALDLF  
 TQDTLEMCTKEMEFCMLFALCYFHAVVAERRKFGAQGWNRSPYFNNGDLTISINVLY  
 NYLEANPKVPWDDLRYLFGEIMYGGHITDDWDRRLCRTYLAEYIRTEMLEGDVLLAPG  
 FQIPPNLDYKGYHEYIDENLPPESPYLYGLHPNAEIGFLTVTSEKLFRTVLEMQPKET  
 DSGAGTGVSREEKVAVLDDILEKIPETFNMAEIMAKAAEKTYPVVVAFQECERMNIL  
 TNEMRRSLKELNLGLKGLTITTDVEDLSTALFYDTVPDTPVARAYPSMMGLAAWYAD  
 LLRIRIRELEAWTDFALPTTVWLAGFFNPQSFLTAIMQSMARKNEWPLDKMCLSEVET  
 KKNREDMTAPPREGSYVYGLFMEGARWDTOTGVIAEARLKELTPAMPVIFIKAI PVDR

METKNIYECPVYKTRIRGPTYVWTFNLKTKEKAAKWILAAVALLQV.
---

Further analysis of the NOV26a protein yielded the following properties shown in Table 26B.

Table 26B. Protein Sequence Properties NOV26a	
PSort analysis:	0.6000 probability located in plasma membrane; 0.4000 probability located in Golgi body; 0.3000 probability located in endoplasmic reticulum (membrane); 0.3000 probability located in microbody (peroxisome)
SignalP analysis:	No Known Signal Sequence Predicted

5 A search of the NOV26a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 26C.

Table 26C. Geneseq Results for NOV26a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV26a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
ABB60101	Drosophila melanogaster polypeptide SEQ ID NO 7095 - Drosophila melanogaster, 4472 aa. [WO200171042-A2, 27-SEP-2001]	1..4454 19..4471	2669/4492 (59%) 3378/4492 (74%)	0.0
AAM78879	Human protein SEQ ID NO 1541 - Homo sapiens, 2143 aa. [WO200157190-A2, 09-AUG-2001]	2314..4455 1..2143	1504/2143 (70%) 1804/2143 (83%)	0.0
AAM79863	Human protein SEQ ID NO 3509 - Homo sapiens, 2127 aa. [WO200157190-A2, 09-AUG-2001]	2254..3929 1..1677	1160/1677 (69%) 1397/1677 (83%)	0.0
AAM79862	Human protein SEQ ID NO 3508 - Homo sapiens, 2127 aa. [WO200157190-A2, 09-AUG-2001]	2254..3929 1..1677	1160/1677 (69%) 1397/1677 (83%)	0.0
AAU74335	Human cytoskeleton-associated protein (CYSKP) #6 - Homo sapiens, 1190 aa. [WO200185942-A2, 15-NOV-2001]	3279..4455 14..1190	1173/1177 (99%) 1175/1177 (99%)	0.0

In a BLAST search of public sequence databases, the NOV26a protein was found to have homology to the proteins shown in the BLASTP data in Table 26D.

Table 26D. Public BLASTP Results for NOV26a				
Protein Accession Number	Protein/Organism/Length	NOV26a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
P23098	Dynein beta chain, ciliary - Tripneustes gratilla (Hawaian sea urchin), 4466 aa.	1..4455 6..4466	3040/4467 (68%) 3658/4467 (81%)	0.0
P39057	Dynein beta chain, ciliary - Anthocidaris crassispina (Sea urchin), 4466 aa.	1..4455 6..4466	3039/4467 (68%) 3657/4467 (81%)	0.0
Q9NYC9	Ciliary dynein heavy chain 9 (Axonemal beta dynein heavy chain 9) - Homo sapiens (Human), 4486 aa.	1..4455 22..4486	2812/4469 (62%) 3518/4469 (77%)	0.0
AAF55834	CG3723-PA - Drosophila melanogaster (Fruit fly), 4496 aa.	1..4454 19..4495	2683/4482 (59%) 3400/4482 (74%)	0.0
Q9VDG0	DHC93AB protein - Drosophila melanogaster (Fruit fly), 4472 aa.	1..4454 19..4471	2669/4492 (59%) 3378/4492 (74%)	0.0

PFam analysis predicts that the NOV26a protein contains the domains shown in the Table 26E.

Table 26E. Domain Analysis of NOV26a			
Pfam Domain	NOV26a Match Region	Identities/ Similarities for the Matched Region	Expect Value
Luteo_ORF3	1022..1055	9/35 (26%) 21/35 (60%)	0.41
Dynein_heavy	3751..4454	434/777 (56%) 674/777 (87%)	0

## 5 Example 27.

The NOV27 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 27A.

Table 27A. NOV27 Sequence Analysis			
	SEQ ID NO: 305	2675 bp	
NOV27a,	CTGTCGTGGTGTGGCTGTGGGACCCGTGAGCAAGCAGCGACGCCAGCGGCGGAGAAC		

CG154465-01 DNA Sequence	<p>CGACGAAAGGTGTCACCACAGTGTGGCAGTGGAGGACAGCACGCTGCAAGTAGTGGT  ACGGGTGCGGCCCCCCCCCTCGGGAGCTGGACAGTCAAGCGCGGCCAGTGGTTTCAG  GTGGTGGACGAGCGGGTGTGGTGTTTAACCTGAGGAGCCCGATGGAGGGTTCCTG  GCCTGAAATGGGGTGGCACCACATGATGGCCCCAAGAAGAAGGGCAAGACCTGACGTT  TGTCTTTGACCGGTCTTTGGCGAGGCGGCCACCCAACAGGACGTGTTCCAGCACACC  ACGCACAGCGTCTTGACAGCTTCTCCAGGGCTACAAGTGTCTAGTGTTCCTTACG  GGGCCACCGGGCTGGGAAGACACACACCATGCTGGGAAGGGAGGGGACCCCGCAT  CATGTACCTGACCACCGTGGAACTGTACAGGCGCTGGAGGCCCGCCAGCAGGAGAAG  CACTTCGAGGTGCTCATCAGCTACCAGGAGGTCTATAATGAACAGATCCATGACCTCC  TGGAGCCCAAGGGGCCCCCTTGCCATCCGCGAGGACCCCGACAAGGGGGTGGTGGTGCA  AGGACTTTCTTTCCACCAGCCAGCCTCAGCCGAGCAGTGTCTGGAGATACTGACGAGG  GGGAACCGTAACCGCAGCAGCACCCCACTGATGCCAACGCGACTTCTTCCGCTCCC  ATGCCATCTTCCAGATCTTTGTGAAGCAGCAGGACCGGGTTCAGGACTGACCCAGGC  TGTCCAGGTGGCCAAGATGAGCCTGATTGACCTGGCTGGCTCAGAGCGGCATCCAGC  ACCCATGCGAAGGGGAGCGGGCTCGGGAGGGGGCCAACATCAACCGCTCTCTGCTGG  CGCTCATCAACGTCCTCAATGCCCTTGGCCGATGCAAGGTAGGCCGCAAGACCCATGT  GCCCTACCGGGACAGCAAACTGACCCGCTGCTCAAAGACTCCCTCGGGGGCAACTGC  CGCACAGTGTATCGCTGCCATCAGCCCCCTCCAGCCTGACCTACGAGGACACGTACA  ACACCCTCAAATATGCCGACCGGGCCAAGGAGATCAGGCTCTCGCTGAAGAGCAATGT  GACCAGCCTGGACTGTACATCAGCCAGTATGCTACCATCTGCCAACAGTCCAGGCCT  GAGGTAGCCGCTCTGAGGAAGAAGCTCCAAGTGTATGAGGGGGGAGGCCAGCCCCAC  CACAGGACCTCCAGGATCTCCAAGTCGGGACCACCACAGAACACCTTCCAGCTC  CCCCTTGCCACCCACCCCTCCAGCCAGCCCTGCACCCAGAGCTCCCTGCAGGGCCT  AGAGCCCTTCAAGAGGAGAGTCTGGGGATGGAGGCCAGGTGGAGAGGGCCATGGAAG  GGAATCTTTCAGACCAGGAGCAGTCCCCAGAGGATGAGGATGAAGGCCAGCTGAGGA  GGTTCCAACCCAGATGCCAGAGCAGAACCCACACATGCACTGCCAGAGTCCCCCTCGC  CTGACCCTGCAGCCCAAGCCAGTCTGTTGGCCACTTCTCAGCACGGGAAGTGGATGGGG  ACCGTTCTAAGCAGTTGGCCCTAAAGGTGCTGTGCGTTGCCAGCGGCAGTACTCCCT  GCTCCAAGCAGCCAACCTCCTGACGCCCCGACATGATCAGAGTTTGAAGCCCTACAG  CAGCTGGTGAAGAGGAAAAAATGAGCCTGGGGCAGAGGCTTGAAGACTTCAGGCC  TGGCCAGGGGGGACCTCTGGCTCAGGAGCTGTGTTTCAAGTCAATCCCTGTGCCGTC  TCTCTCTGCCAGAGCCTCCAGGATACACTGGCCCTGTGACCCGACTATGGCGAGG  CGACTGAGTGGCCCCCTGCACACCCCTGGGAATCCCGCCTGGACCCCACTGCACCCAG  CCCAGGGGTCCCGATGGCCCATGGAGAAGAAGAGGAGGAGACCAAGCGCCTTGGAGGC  AGACAGTCCCATGGCCCCAAGCGGGGCACCAAGCGCCAGCGCCAGTCTTCTGCC  TGCCTAAGGAGAGGGTCTCTGCTGACACCCAACCTTCACAGGGGCCAGCACCCCA  AAGGAGAAAGGGCCTCCTCCCCCTGCCATTCCCCTCGCGTTTGCCAGCCACAGTCAT  CAAAAGCCGGGTGCCCCCTGGGCCCTTCCGCCATGCAGAAGTGTCTCCACCCGCTGGCT  CTGCCCCACTCGAGACCTCAATGCCACCTTTGATCTCTCTGAGGAGCTCCCTCAAAGC  CAGTTTCCATGAATGCATTGGCTGGGACAAAATACCCAGGAGCTGAGCAGGCTGGA  CCAGCCCTTCATCCCAGGGACCTGTGCCCTGTTTACCATGAAGGGCCCCAAGCCA  ACATCTTCCCTCCCTGGGACCTCTGCCTGCAAGAAGAAGCGCTTGCAGTTTCTCAG  TCTCCATGGCCGAGCCGATCGCCCGCTCCCGAGCAGCACTTGAAGAGGCCAGC  TGGGCCCTTGTACTCCAGGTGACTGGCACTAGGGACAGGGATAGCTTGGGCATGG  AGGCCGATGAAGACAAGAAGGAGGAGGGGACGGGGAGCTGAGACCCAGAAGAAAGGAG  GGCCTAG</p>
	<div>ORF Start: ATG at 82</div> <div>ORF Stop: TAG at 2584</div>
	<div>SEQ ID NO: 306</div> <div>834 aa</div> <div>MW at 91153.5kD</div>
NOV27a, CG154465-01 Protein Sequence	<p>MAVEDSTLQVVVRVPPTPRELDSQRRPVVQVVDERVLVFNPEEPDGGFPGLKWGGTH  DGPKKKGKDLTFVDRVFGAATQQDVFOHTTHSVLDSFLQYNCVSFAYGATGAGKT  HTMLGREGDPGIMYLTVELYRRLEARQQEKHFVLSIQEYVNEQIHDLLEPKGPLA  IREDPDKGVVQGLSFHQPSAEQLEILTRGNRNRTQHPTDANATSSRSHAFQIFV  KQQDRVPLTQAVQVAKMSLIDLAGSERASSTHAKGERLREGANINRSLALINVLNA  LADAKVGRKTHVPYRDSKLRLLKDSLGGNCRTVMIAAISPSLLTYEDTYNTLKYADR  AKEIRLSLKSNTSLDCHISQYATICQQLQAEVAALRKQLQVYEGGQPPQDLPGSP  KSGPPPEHLPSPLPPHPSQPCTPELPAGPRALQEESLGMEAQVERAMEGNSDQEQ  SPEDEDEGPAEEVPTQMPEQNPTHALPESPRLTLPKPVVGHFSARELDGDRSKQLAL  KVLCAQRQYSLQAANLLTPDMI TEFETLQQLVQEEKIEPGAELRTSGLARGAPLA  QELCSESI PVPSPLCPEPPGYTGPTVTRTMARRLSGPLHTLGI PPGPNCTPAQGSRWPM</p>

	EKKRRRPSALEADSPMAPKRGTKRQSQSFLPCLRRGSLPDTQPSQGPSTPKGERASSP CHSPRVCPATVIKSRVPLGPSAMQNCSTPLALPTRDLNATFDLSEPPSPKPSFHECIG WDKIPQELSRLDQPFIPRAPVPLFTMKGPKPTSSLPGTSACKKKRVASSSVSHGRSRI ARLPSTLKRPAGPLVLPGDWH
--	---

Further analysis of the NOV27a protein yielded the following properties shown in Table 27B.

Table 27B. Protein Sequence Properties NOV27a	
PSort analysis:	0.7000 probability located in nucleus; 0.4267 probability located in mitochondrial matrix space; 0.3000 probability located in microbody (peroxisome); 0.1042 probability located in mitochondrial inner membrane
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV27a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 27C.

5

Table 27C. Geneseq Results for NOV27a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV27a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
ABB07410	Human kinesin motor protein, HsKip3A - Homo sapiens, 864 aa. [WO200196593-A2, 20-DEC-2001]	1..830 1..829	828/830 (99%) 828/830 (99%)	0.0
ABB07412	Amino acid sequence of Kip3A fragment used in ATPase assay - Homo sapiens, 383 aa. [WO200196593-A2, 20-DEC-2001]	1..360 1..359	354/360 (98%) 355/360 (98%)	0.0
ABB07411	Human HsKip3A motor domain fragment - Homo sapiens, 338 aa. [WO200196593-A2, 20-DEC-2001]	5..343 1..338	338/339 (99%) 338/339 (99%)	0.0
AAU76967	Novel human kinesin motor protein, HsKip3d insertion mutant - Homo sapiens, 905 aa. [WO200212268-A1, 14-FEB-2002]	8..392 12..402	231/391 (59%) 298/391 (76%)	e-130
AAU76957	Novel human kinesin motor protein, HsKip3d - Homo sapiens, 898 aa. [WO200212268-A1, 14-FEB-2002]	8..392 12..395	231/385 (60%) 297/385 (77%)	e-130

In a BLAST search of public sequence databases, the NOV27a protein was found to have homology to the proteins shown in the BLASTP data in Table 27D.

Table 27D. Public BLASTP Results for NOV27a				
Protein Accession Number	Protein/Organism/Length	NOV27a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q91WD7	Similar to hypothetical protein DKFZp434G2226 - Mus musculus (Mouse), 886 aa.	8..392 12..395	233/385 (60%) 296/385 (76%)	e-131
BAB93508	OK/SW-CL.108 - Homo sapiens (Human), 898 aa.	8..392 12..395	231/385 (60%) 297/385 (77%)	e-129
Q9H0F3	Hypothetical 102.3 kDa protein - Homo sapiens (Human), 898 aa.	8..392 12..395	231/385 (60%) 297/385 (77%)	e-129
Q9VSW5	KLP67A protein (RE52076p) - Drosophila melanogaster (Fruit fly), 814 aa.	4..452 5..434	213/451 (47%) 283/451 (62%)	3e-99
P91945	Kinesin like protein 67A - Drosophila melanogaster (Fruit fly), 814 aa.	4..452 5..434	213/451 (47%) 283/451 (62%)	3e-99

PFam analysis predicts that the NOV27a protein contains the domains shown in the Table 27E.

Table 27E. Domain Analysis of NOV27a			
Pfam Domain	NOV27a Match Region	Identities/ Similarities for the Matched Region	Expect Value
kinesin	13..388	158/435 (36%) 281/435 (65%)	2.3e-114

### Example 28.

The NOV28 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 28A.

5

Table 28A. NOV28 Sequence Analysis			
	SEQ ID NO: 307	1872 bp	
NOV28a, CG154492-01 DNA Sequence	CGCGGGCGGCTGGCGTCGGGAAAAGTACAGTAAAAAGTCCGAGTGCAGCCGCCGGGCGCA GGATGGGATCCGGCTCCTCCAGCTACCGGCCAAGGCCATCTACCTGGACATCGATGG ACGCATTCAGAAGGTAATCTTCAGCAAGTACTGCAACTCCAGCGACATCATGGACCTG TTCTGCATCGCCACCGGCCTGCCTCGGAACACGACCATCTCCCTGCTGACCACCGACG ACGCCATGGTCTCCATCGACCCCAACATGCCCGCGAATTCAGAACGCACTCCGTACAA AGTGAGACCTGTGGCCATCAAGCAACTCTCCGCTGGTGTGCGAGGACAGAGAACCA AGCCGTGGCCAGTCTGCTGAGAGACCACTGAGGGACAGACGGGTTGTGGGCCTGGAGC AGCCCCGAGGGAAGGAGCATTTGAAAGTGGACAGGTAGAGCCCAGGCCAGAGAGCC CCAGGGCTGCTACCAGGAAGGCCAGCGCATCCCTCCAGAGAGAGAAGAATTAATCCAG		

	AGCGTGCTGGCGCAGGTTGCAGAGCAGTTCTCAAGAGCATTCAAATCAATGAAGTGA AAGCTGAAGTTGCAAATCACTTGGCTGTCTAGAGAAACGCGTGAATTTGAAGGACT AAAAGTGGTGGAGATTGAGAAATGCAAGAGTGACATTAAGAAGATGAGGGAGGAGCTG GCGGCCAGAAGCAGCAGACCAACTGCCCTGTAAAGTACAGTTTTTGGATAACCA AGAAAGTTGACTCCTCGACGCGATGTTCCCACTTACCCCAAGTACCTGCTCTCCAGA GACCATCGAGGCCCTGCGGAAGCCGACCTTTGACGTCTGGCTTTGGGAGCCCAATGAG ATGCTGAGCTGCCTGGAGCACATGTACCACGACCTCGGGCTGGTCAGGGACTTCAGCA TCAACCCGTGTCAACCTCAGGAGGTGGCTGTTCTGTGTCCACGACAACCTACAGAAACAA CCCCTTCCACAACCTCCGGCACTGCTTCTGCGTGGCCAGATGATGTACAGCATGGTC TGGCTCTGCAGTCTCCAGGAGAAGTTCTCACAACCGGATATCCTGATCCTAATGACAG CGGCCATCTGCCACGATCTGGACCATCCCGGCTACAACAACCGTACCAGATCAATGC CCGCACAGAGCTGGCGGTCCGCTACAATGACATCTCACCGCTGGAGAACCACCACTGC GCCGTGGCCTTCCAGATCCTCGCCGAGCCTGAGTGCAACATCTTCTCCAACATCCAC CTGATGGGTTCAAGCAGATCCGACAGGGAATGATCACATTAATCTTGGCCACTGACAT GGCAAGACATGCAGAAATTATGGATTCTTTCAAAGAGAAAATGGAGAATTTTGACTAC AGCAACGAGGAGCACATGACCTCAGCGACCGTGAGAAGTCAGAAGGCCCTTCTGTGG CACCGTTCATGGACCGAGACAAAGTGACCAAGGCCACAGCCAGATTGGGTTTCATCAA GTTTGTCTGATCCCAATGTTTGAACAGTGACCAAGCTCTTCCCATGTTTGAGGAG ATCATGCTGCAGCCACTTTGGGAATCCCGAGATCGCTACGAGGAGCTGAAGCGGATAG ATGACGCCATGAAAGAGTTACAGAAGAAGACTGACAGCTTGACGTCTGGGGCCACCGA GAAGTCCAGAGAGAGAAGCAGAGATGTGAAAAACAGTGAAGGAGACTGTGCCCTGAGGA AAGCGGGGGCGTGGCTGCAGTTCTGGACGGGCTGGCCGAGCTGCGCGGGATCCTTGT GCAGGGAAGAGCTGCCCTGGGCACCTGGCACCACAAGACCATGTTTTCTAAGAACCAT TTTGTTCACCTGATACA		
	ORF Start: ATG at 61	ORF Stop: TGA at 1735	
	SEQ ID NO: 308	558 aa	MW. at 64319.9kD
NOV28a, CG154492-01 Protein Sequence	MGSGSSSYRPKATYLDIDGRIQKVIKSKYCNSSDIMDLFCIATGLPRNTTISLLTDD AMVSIDPTMPANSERTPYKVRPVAIKQLSAGVEDKRTTSRQSAERPLRDRRVVGLAQ PRREGAFESGQVEPRPREPQGCYQEGQRIPPERELIQSVLAQVAEQFSAFKINELK AEVANHLAVLEKRVELEGLKVVEIEKCKSDIKMREELAARSRTNCPCKYSFLDNHK KLTPRRDVPTYPKYLLSPETIEALRKPTFDVWLWEPNEMLSLEHMYHDLGLVRDFSI NPVTLRRWLFVHDNYRNNPFHNRHCFVQOMYSMVWLCSLQEFQSDILILMTA AICHDLDPGYNNTYQINARTELAVERYNDISPLENHHCVAFAQILAEPECNIFSNIPP DGFKQIRQGMITLLLATDMARHAEIMDSFKEKMFNFDYSNEHMTLSREKSEGLPVA PFMDRDKVTKATAQIGFIKFLVLI PMFETVTKLFPVVEIMLQPLWESRDYBELKRID DAMKELQKKTDSLTSGLATEKSRERSRDVKNSEGDCA		
	SEQ ID NO: 309	1653 bp	
NOV28b, CG154492-02 DNA Sequence	CGGGAAGTACAGTAAAAAGTCCGAGTGACGCCACCGGGCGCAGGATGGGGTCCGGCT CCTCCGGCTACCGGCCCAAGGCCATCTACCTGGACATCGATGGACGATTGAGAAGGT AATCTTCAGCAAGTACTGCAACTCCAGCGACATCATGGACCTGTTCTGCATCGCCACC GGCCTGCCTCGGAACACGACCATCTCCCTGCTGACCACCGACGACGCAATGGTCTCCA TCGACCCCAACATGCCCGCAATTGAGAACGCACTCCGTACAAAGTGAGACCTGTGGC CATCAAGCAACTCTCCGAGAGAGAAGAATTAATCCAGAGCGTGCTGGCGCAGGTTGCA GAGCAGTTCTCAAGAGCATTCAAATCAATGAACTGAAAGCTGAAAGTGCAATCACT TGGCTGTCTAGAGAAACCGGTGGAATTGGAAGGACTAAAAGTGGTGGAGATTGAGAA ATGCAAGAGTGACATTAAGAAGATGAGGAGGAGCTGGCGGCCAGAAGCAGCAGGACC AACTGCCCCGTGAAGTACAGTTTTTTGGATAACCAACAAGAGTTGACTCCTCGACGCG ATGTTCCCACTTACCCCAAGTACCTGCTCTCTCCAGAGACCATCGAGGCCCTGCGGAA GCCGACCTTTGACGTCTGGCTTTGGGAGCCCAATGAGATGCTGAGCTGCCTGGAGCAC ATGTACCACGACCTCGGGCTGGTCAGGGACTTCAGCATCAACCCCTGTACCCCTCAGGA GGTGGCTGTTCTGTGTCCACGACAACCTACAGAAACAACCCCTTCCACAACCTCCGGCA CTGCTTCTGCGTGGCCAGATGATGTACAGCATGGTCTGGCTCTGAGTCTCCAGGAG AAGTTCTCACAACCGGATATCCTGATCCTAATGACAGCGGCCATCTGCCACGATCTGG ACCATCCCGGCTACAACAACACGTACCAGATCAATGCCCGCACAGAGCTGGCGGTCCG CTACAATGACATCTCACCGCTGGAGAACCACCACTGCGCGTGGCTTCCAGATCCTC GCCGAGCCTGAGTGCAACATCTTCTCCAACATCCCACCTGATGGGTTCAAGCAGATCC GACAGGGAATGATCACATTAATCTTGGCCACTGACATGGCAAGACATGCAGAAATTAT GGATTCTTTCAAAGGGAAAATGGAGAATTTTGACTACAGCAACGAGGAGCACATGACC		

	CTGCTGAAGATGATTTTGATAAAATGCTGTGATATCTCTAACGAGGTCGCTCCAATGG AAGTCGCAGAGCCTTGGGTGAGACTGTTTATTAGAGGAATATTTATGCAGAGCGACCG TGAGAAGTCAGAAGGCCTTCTGTGGCACCCTTCATGGACCGAGACAAAGTGACCAAG GCCACAGCCCAGATTGGGTTCATCAAGTTTGTCTCTGATCCCAATGTTTGAACAGTGGA CCAAGCTCTTCCCCATGGTTGAGGAGATCATGCTGCAGCCACTTTGGGAATCCCGAGA TCGCTACGAGGAGCTGAAGCGGATAGATGACGCCATGAAAGAGTTACAGAAGAAGACT GACAGCTTGACGTCTGGGGCCACCGAGAAGTCCAGAGAGAGAAGCAGAGATGTGAAAA ACAGTGAAGGAGACTGTGCCTGAGGAAAG		
	ORF Start: ATG at 46		ORF Stop: TGA at 1645
	SEQ ID NO: 310	533 aa	MW at 61606.3kD
NOV28b, CG154492-02 Protein Sequence	MGSGSSGYRPKAIYLDIDGRIQKVIFSKYCNSSDIMDLFCIATGLPRNTTISLLTTDD AMVSIADPTMPANSERTPYKVRPVAIKQLSEREELIQSVLAQVAEQFSRAFKINELKAE VANHLAVLEKRVELEGLKVVEIEKCKSDIKKMREELAARSSRTNCPCKYSFLDNHKKL TPRRDVPTYPKYLLSPETIEALRKPTFDVWLWEPNEMLSCLEHMYHDLGLVRDFSINP VTLRRWLCVHDNYRNNPFHFRHCFCVAQMMYSMVWLCSLQEKFSQTDILILMTAAI CHDLDPGYNNYQINARTELAVRYNDISPLENHHCAVAFQILAEPECNIFSNIPPDG FKQIRQGMITLILATDMARHAEIMDSFKGKMENFDYSNEEHMTLLKMILLIKCCDISNE VRPMEVAEPWVDCLEEYFMQSDREKSEGLPVAPFMDRDKVTKATAQIGFIKFVLI PM FETVTKLFPMBEIMLQPLWESDRDYEBELKRIDDAMKELQKKTDSLTSGATEKSRERS RDVKNSEGDCA		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 28B.

Table 28B. Comparison of NOV28a against NOV28b.		
Protein Sequence	NOV28a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV28b	1..558	461/593 (77%)
	1..533	470/593 (78%)

Further analysis of the NOV28a protein yielded the following properties shown in Table 28C.

Table 28C. Protein Sequence Properties NOV28a	
PSort analysis:	0.7600 probability located in nucleus; 0.1000 probability located in mitochondrial matrix space; 0.1000 probability located in lysosome (lumen); 0.1000 probability located in plasma membrane
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV28a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 28D.

Table 28D. Geneseq Results for NOV28a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV28a Residues/	Identities/ Similarities for	Expect Value

		<b>Match Residues</b>	<b>the Matched Region</b>	
ABG61846	Prostate cancer-associated protein #47 - Mammalia, 593 aa. [WO200230268-A2, 18-APR-2002]	1..558 1..593	558/593 (94%) 558/593 (94%)	0.0
AAV28561	Cyclic-GMP specific phosphodiesterase (PDE9A) - Homo sapiens, 593 aa. [WO9929873-A1, 17-JUN-1999]	1..558 1..593	558/593 (94%) 558/593 (94%)	0.0
AAV39285	Phosphodiesterase 10 (PDE10) clone FB68.2 - Homo sapiens, 580 aa. [WO9942596-A2, 26-AUG-1999]	14..558 1..580	544/580 (93%) 544/580 (93%)	0.0
AAV39284	Phosphodiesterase 10 (PDE10) clone FB76.2 - Homo sapiens, 533 aa. [WO9942596-A2, 26-AUG-1999]	1..558 1..533	463/593 (78%) 472/593 (79%)	0.0
AAB92673	Human protein sequence SEQ ID NO:11043 - Homo sapiens, 474 aa. [EP1074617-A2, 07-FEB-2001]	148..558 29..474	411/446 (92%) 411/446 (92%)	0.0

In a BLAST search of public sequence databases, the NOV28a protein was found to have homology to the proteins shown in the BLASTP data in Table 28E.

<b>Table 28E. Public BLASTP Results for NOV28a</b>				
<b>Protein Accession Number</b>	<b>Protein/Organism/Length</b>	<b>NOV28a Residues/ Match Residues</b>	<b>Identities/ Similarities for the Matched Portion</b>	<b>Expect Value</b>
O76083	High-affinity cGMP-specific 3',5'-cyclic phosphodiesterase 9A (EC 3.1.4.17) - Homo sapiens (Human), 593 aa.	1..558 1..593	558/593 (94%) 558/593 (94%)	0.0
AAH09047	Similar to phosphodiesterase 9A - Homo sapiens (Human), 533 aa.	1..558 1..533	463/593 (78%) 472/593 (79%)	0.0
O70628	High-affinity cGMP-specific 3',5'-cyclic phosphodiesterase 9A (EC 3.1.4.17) - Mus musculus (Mouse), 534 aa.	1..555 1..529	423/590 (71%) 456/590 (76%)	0.0
Q8QZV1	cGMP phosphodiesterase - Rattus norvegicus (Rat), 534 aa.	1..554 1..528	420/589 (71%) 457/589 (77%)	0.0
AAF48205	CG32648-PA - Drosophila melanogaster (Fruit fly), 963 aa.	249..549 48..380	152/336 (45%) 199/336 (58%)	4e-78

Pfam analysis predicts that the NOV28a protein contains the domains shown in the Table 28F.

Table 28F. Domain Analysis of NOV28a			
Pfam Domain	NOV28a Match Region	Identities/ Similarities for the Matched Region	Expect Value
PDEase	311..440	55/133 (41%) 90/133 (68%)	9.8e-52
PDEase	454..498	14/47 (30%) 33/47 (70%)	1.1e-08

### Example 29.

5 The NOV29 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 29A.

Table 29A. NOV29 Sequence Analysis			
	SEQ ID NO: 311	13332 bp	
NOV29a, CG154509-01 DNA Sequence	CTCCGGACTGGTTTCTTCTTCCTTCCCCCTTCCCCCAACTTCCCTCCACCCCTTCCAA TCATGGCGAACGGGACTGCGGACGTTTCGGAACTCTTCATCTTCACTACTACCCAGAA TTACTTCGGGTTGATGTCGAACTCTGGGATCAGCCACTGTTGTGCAACTGTCTTGAA ATCAACAACCTCTTGGATGACGGCAACCAGATGCTCCTCAGGGTGACGATCCGACG CAGGAATCTCCTTTTCCAACACGATTGAGTTTGGTGACACAAAAGATAAAGTGCTGGT GTTTTTCAAGCTGCGACCTGAAGTAATTACTGATGAGAATCTACATGATAACATTCTT GTTTCATCTATGTTAGAGTCACCTATTAGTTCTCTTTACCAAGCAGTACGGCAAGTAT TCGCACCAATGTTGTTAAAGGATCAGGAATGGAGCAGAACTTTGATCCCAAACTTCA GAATCTTTTGAGTGAAC TAGAGCTGGGTTGGGTATAGTTCTACGAAGATCAGACACT AACTTAACAAAATTGAAATTTAAGGAAGATGACACACGAGGTATCCTTACACCAAGCG ATGAGTCCAGTTTGGATAGAACAGCTCACCGTGGAAATAAACAGATTAGTAAAGA AAGAGCCAATTATTTTAAAGAATTATTTGAAACAATTGCAAGAGAGTTTATAACTTG GACAGTCTATCCTTACTAGAAGTTGTTGACTTGGTGGAGACTACTCAGGATGTTGTAG ATGATGTGTGGAGACAAACAGAACATGATCATTATCCTGAGTCACGAATGTTGCATCT CTTAGACATCATAGGTGGTTTCAATTTGGAAGGTTTGTTCAGAAAAAGTTGGGAACCTTG AACCTGTGGGAAGATCCTTATTATCTTGTGAAAGAAAGCTGAAAGCTGGTATTTCAA TTTGTGAACAGTGGGTGATAGTCTGTAATCATCTAACAGGTGAGGTGGCAGCGCTA TGTTCTCATCCATGGAAAAATGAAAAATATTTCCAGAAACACTTGACAAACTTGGC AAACGCCTTGAAGAGGTCCTTGGCTATTAGAACAATTCATGAGAAGTTCTCTATTTTC TACCTGCCAGTGAAGAGAAAAATCATATGCCTCACTCGAGTATTTGAACCTTTTACTGG CCTGAATCCTGTGCAATATAATCCATATACTGAGCCCTTGTGGAAAGCTGCGGTGTCT CAATATGAAAAGATTATTGCACCTGCGGAACAAAAATAGCAGGAAAATTGAAAAATT ATATTTAGAAAATTCAAGACAGTCCACAGCAGCTTCTTCAAGCATTCCTGAAATATAA AGAGTTGGTAAAGCGTCCAATAAGCAAAGAATTGATGTTAGAAAAGAGAACTTTA CTGGCAAGACTTGTGGACTCAATTAAAGATTTTCGATTAGACTTTGAGAATCGGTGCC GAGGAATTCCTGGTGATGTCATCTGGACCACTTTCTGGCAAAAATCTTTCAGAAGTTGT CAACAGTATAGTTTGGGTTCCGCAAGTTGGAATTGAAGGTAGATGATACTATCAAGACT GCAGAGGCTCTTTTATCTGACTTGCCAGGATTTTCGATGTTTCCATCAAAGTGCCAAAG ATCTCTTAGACCAGCTTAAACTATATGAACAGGAACAATTTGATGATTGGTCCAGGGA TATTCATCAGGTTTATCTGATTCCAGATCTGGTTTGTGTATTGAGGCTAGTAGTCGA ATTATGGAATTGGATTCTAATGATGGATTACTAAAAGTGCATTATTCAGATCGTTTGG TGATTCTTCTGAGAGAAGTTCGTGAGCTCTCTGCACTTGGCTTTTGTATTCCTGCCAA AATACAGCAAGTTGCAACATTGCACAGAAATCTGCAAGCAAGCAATTATCTTAAA CAAGTGGCACATTTTATAATTCTATTGATCAACAAATGATTCAAAGTCAGAGGCCAA		

TGATGTTACAATCTGCCTTAGCATTTGAACAGATAATTAAGAATTCAAAAGCAGGAAG  
 TGGAGGGAAATCACAGATAACTTGGGATAATCCTAAAGAATTAGAAGGCTATATCCAA  
 AAACCTCCAAAATGCTGCTGAACGGCTTGCCACTGAAAATAGAAAACCTGAGAAAATGGC  
 ACACCTACATTTTGTGAAAAGGTGGTTGTTCTTATGAATATTGATCTGCTTCGGCAGCA  
 ACAGCGCTGGAAGATGGATTACAAGAATTGAGAACTGGCTTAGCAACTGTAGAAGCA  
 CAGGGATTCCAAGCAAGTGACATGCATGCATGGAAACAACACTGGAATCATCAACTGT  
 ACAAAGCTCTGGAGCATCAGTACCAGATGGGCTTAGAAGCACTTAATGAGAATTTGCC  
 AGAAATAAATATAGACTTAACCTACAACAGGGACGATTACAATTCAGGCCCTTTT  
 GAAGAAATCCGGGCTAAATATTATAGAGAAATGAAGAGATTTCATCGGCATTCCAAATC  
 AGTTTAAGGGAGTGGGTGAGGCCAGGAGCATTAAATCTATTTTTCTATTATGATTGA  
 TAGAAAATGCAAGTGGATTTTTGTACGATTTTTCAGCAAAGCTGAACATCTGTTTAGAAGA  
 TTGTCAGCTGTTTTACACCAACATAAGGAATGGATTGTAATTGGGCAAGTTGATATGG  
 AAGCTCTGGTGGAAAAGCATCTTTTACTGTACATGATTGGGAGAAAAATTTTAAAGC  
 ATTAATAAATAAAGGGGAAAGAAGTAGAACGACTTCCAAGTCTGTCAAGGTAGATTGT  
 TTAAATATTAATTGCAACCCTGTGAAGACTGTGATTGATGATCTCATCCAGAAGTTAT  
 TTGATCTGCTTGTCTTTCTTTGAAGAAGTCCATACAGGCTCATTTACATGAAATTGA  
 TACATTTGTTACTGAGGCTATGGAAGTCTTAACAATTATGCCCCAGTCTGTGGAAGAA  
 ATTGGTGATGCAAACTACAATATAGTAAGTTACAAGAACGGAAGCCAGAGATTTTGC  
 CCTTATTTCAAGAAGCTGAAGACAAAACAGACTTTTACGAACTGTGGCTGGTGGAGG  
 TTTAGAAAACATTAAGTAATTTGAAAGCCAAGTGGGATAAATTTGAGTTAATGATGGAA  
 AGTCACCAACTTATGATTAAAGACCAGATTGAAGTGATGAAAGGAAATGTGAAATCAC  
 GTCTTCAGATCTATTATCAAGAACTGGAAAAATTTAAAGCTCGTTGGGACCAACTAAA  
 GCCTGGTGATGATGTTATTGAAACTGGCCAACATAACTCTTGATAAAAGTGCAAAG  
 TTAATAAAAGAGAAAAAAATTGAGTTTGATGATCTTGAAGTCACAAGAAAAAGCTGG  
 TTGATGATTGCCATCATTTTAGACTGGAAGAGCCTAATTTCTCCCTGGCAAGTAGTAT  
 CTCTAAAGATATCGAGAGCTGTGCCCAAATTTGGGCCCTTTTATGAAGAGTTTCAACAA  
 GGATTTTCAGGAAATGGCCAATGAAGACTGGATCACTTTTCGGACTAAGACATACCTGT  
 TTGAGGAATTTTGTGTAACCTGGCATGACAGATTAAGGAAGGTTGAAGAACATTCAGT  
 GATGACAGTGAAATTACAATCAGAGGTTGACAAATATAAAATCGTAATTCCTATCTTG  
 AAATATGTGAGAGGGGAGCATCTTCTCCAGATCACTGGCTTGACCTTTTTCGTCTCC  
 TTGGACTTCTAGGGGACTAGTCTAGAGAACTACTGTTTGGTGATTGCTCAGAGT  
 AGCTGATACAATTGTAGCCAAAGCTGCCGACCTTAAAGATTTAAATAGTCGGGCACAA  
 GGTGAAGTTACAATCAGAGAAGCTTTACGTGAACCTGATCTTTGGGGAGTTGGAGCAG  
 TGTTTACATTAATTGATTATGAAGACAGCCAAAGTCGAACATGAAGCTGATTAAAGA  
 CTGGAAAGATATAGTAAATCAGGTTGGAGATAATAGATGCCTTCTCCATCCTTAAAG  
 GATTCTCCTTATTATAAAGGATTTGAAGATAAAGTATCAATTTGGGAAAGAAAACCTTG  
 CAGAGTTAGATGAATACCTGCAGAATTTAAATCATATTCAGAGAAAGTGGGTGTATTT  
 GGAACCCATTTTTCGGCCGTGGAGCATTGCCAAAAGAACAGACACGTTTCAACAGAGTT  
 GATGAAGATTTTAGATCAATAATGACTGATATCAAGAAAGACAATAGAGTCACAACAT  
 TAACTACTCATGCTGGAATAAGAAATTCTCTACTAACAATACTTGATCAGCTTCAAAG  
 ATGTGAGAGATCATTAATGAATTTTGGAGGAAAAACGCTCAGCATTTCCCAAGATTT  
 TATTTTATTGGTGATGATGACTTATTAGAAATATTGGGCCAGTCTACCAACCCATCAG  
 TGATTGAGTCTACCTGAAGAAGCTTTTGTGCTGGTATTACAGTGTTTGCCTTTGATGA  
 GAAATCAAAACATATAACTGCAATGAAATCTTTAGAGGGAGAAAGTTGTACCTTTTAAA  
 AATAAAGTTCTCTATCAAATAATGTAGAGACATGGTTGAATGATTGGCCTTAGAAA  
 TGAAGAAAACCTTTGGAACAGTTGTTGAAGGAATGTGTTACTACTGGGCGAAGTTCTCA  
 AGGTGCAGTTGACCCATCTCTGTTCCCTTCACAGATTTTATGCTTGGCGGAGCAGATT  
 AAATTCAGTGAAGATGTAGAAAATGCTATTAAAGATCATAGTCTTCATCAGATTGAAA  
 CACAACTGGTGAATAAGTTAGAGCAATATACTAACATTGATACAAAGTTCTGAGGATCC  
 AGGGAATACTGAATCGGGCATCCTGGAGCTTAAACTTAAAGCCCTAATCTTGACATT  
 ATCCATAATATTGATGTGGTAAAGCAGTTAAACCAATTCAGGTTTATACAACCTGAAG  
 ACTGGGCTTGGAAAAACAACCTTAGATTCTATATGAAAAGTGATCATACATGTTGTGT  
 TCAAATGGTGGATTCTGAATTTTCACTACTTATGAATATCAGGGTAATGCTTCCAAA  
 CTGGTTTATACCTCACTGACAGACAAGTGCTACTTAACCTCACTCAAGCCATGAAGA  
 TGGGACTTGGAGGAAATCCTTATGGACCAGCTGGAACCTGGGAAAACGGAATCAGTAAA  
 GGCTTTAGGTGGACTTCTTGAAGACAAGTTTGTAGTCTTTAATTGTGATGAGGGCATC  
 GATGTGAAGTCAATGGGACGAATATTTGTGGTTTGGTGAAGTGTGGGCCTGGGGTT  
 GTTTGATGAATTTAATAGGCTGGAAGAATCTGTACTGTCAGCAGTTTCTATGCAAT  
 CCAGACAATTCAGATGCTTTGAAGAATCATAGAAGTGTATGTAAGTGTCTGGCAAG  
 GAGGTAGAAGTAAATTTCTAATTTCTGGAATTTTATCACTATGAATCTGCTGGAAAAAG  
 GTTATGGAGGAAGACAAAACCTGCCTGATAATCTTAAACAGCTTTTCAGGCCGCTAGC

TATGTCTCATCCAGACAATGAGCTTATTGCAGAAGTTATTCTCTATTTCGGAAGGCTTT  
 AAAGACGCTAAAGTATTGAGCAGAAAATTGGTAGCTATTTTCAATCTATCTAGGGAAC  
 TTTTGACACCTCAGCAACATTATGATTGGGGTTTGAGAGCTTTGAAGACAGTTCTGAG  
 AGGAAGTGGAATCTCCTTAGACAGCTAAACAAAAGTGGCACTACACAGAATGCTAAT  
 GAAAGTCATATTGTGGTACAAGCACTGAGGCTTAATACAATGTCAAAGTTTACGTTTA  
 CTGATTGCACCCGGTTTGATGCACTGATAAAAAGATGTCTTTCCGGGAATTGAATTGAA  
 AGAAGTGGAATATGATGAACCTAAGTGTCTGCATTAAAGCAGGTCTTTGAAGAGGCCAAT  
 TATGAAATTATACCCAATCAGATCAAAAAGGCTTTAGAATTGTATGAACAGTTATGCC  
 AGAGGATGGGAGTTGTTATTGTTGCTCAAGTGGTGTGGAAAATCAACGCTTTGGAG  
 AATGTTAAGGGCTGCGCTTTGTAAAACCTGGCAAAGTAGTGAACAATATACTATGAAT  
 CCCAAAGCTATGCCCTCGATATCAATTATTAGGCCATATTGACATGGACACAAGAGAAT  
 GGTCTGATGGTGTTTTGACAAATAGTGTCTCGTCAAGTGGTTCGGGAACCTCAAGATGT  
 CAGCTCATGGATAATCTGTGATGGTATATTGACCCTGAATGGATAGAATCTCTGAAT  
 TCTGTTCTGGATGATAATCGACTGCTGACTATGCCAGTGGAGAAAGGATTCAGTTTG  
 GCCCAAATGTTAATTTGTATTGAAACTCATGATTTAAGTTGTGCATCACCAGCCAC  
 AATATCTAGAATGGGAATGATCTTTCTTAGTGATGAAGAGACAGATCTTAATCTCTG  
 ATAAAATCTTGGTTGAGGAATCAGCCTGCTGAATATAGAAAATCTTGAAGAAATTGGA  
 TTGGAGATTATTTTGAAGAGGCTTTACAATGGGTTCTAAAGCAGAATGACTATGTGGT  
 AGAAACAAGTTTGGTTGGGACTGTGATGAATGGTTGTGCATCTACATGGTTGCAGA  
 GATCATGACGAATTCATTATTAATCTCATAAGGGGACTTGGTGGAAATCTGAATATGA  
 AGTCACGTTTGGAAATTTACCAAAGAGGTTTTTCATTGGGCACGAGAATCTCCTCCAGA  
 CTTTCACAAACCTATGGATACCTACTATGACTCTACTAGGGGTTCGATTAGCAACATAT  
 GTGCTTAAGAAGCCAGAAGACTTGACTGCTGATGATTTTCAGTAACGGCTTAACCTTTC  
 CAGTCATTTCAGACTCCTGACATGCAACGAGGTCTAGATTATTTCAAACCATGGTTAAG  
 TTCTGATACTAAACAGCCCTTTATTCTGGTAGGACCAGAAGGATGTGGCAAAGGGATG  
 CTGCTCAGGTACGCATTTTCACAACTCCGGTCCACTCAAATTGCTACAGTTCACTGTA  
 GTGCACAAACCACTTCTCGACATCTCCTGCAGAACTGAGCCAGACTTGCATGGTAAT  
 CAGTACTAATACTGGTCTGTATACAGACCAAAAGACTGTGAAAGACTTGTCTGTAC  
 TTAAGATATCAACCTACCTAACTTGATAAATGGGGGACCAGTACTTTGGTAGCAT  
 TCCTACAACAGGTATTGACGTATCAAGGATTTTATGATGAAAATTTGGAATGGGTTGG  
 TCTAGAAAATATTCAAATTTGGCTTCTATGTGCTGAGTGGAGGAAGACTGGGAAGACAT  
 AAACCTACTACCAGATTTACTTCCATCGTTCTTTGTTCTATAGATTACCCAGAAA  
 GAGAGCAGTTACAAACGATTATGAGCATAATTTGAACCAAGTTCTACATAAAAATCT  
 GAAGAATCATTCTATTTGGGGTCTTCATCAAAAATTTATCTTTTAGCAGGATCTATG  
 GTACAAGTGTATGAACAGGTAGATATGCATCAGGTGCGAGCCAAATTTACAGTTGATG  
 ATTATAGTCACTATTTCTTTACTCCTTGCACTTCTTACCCAATGGGTTCTTGGCTTATT  
 TAGATATGATTTAGAAGGAGGATCCTCAAACCATCCACTAGATTATGTGTAGAAAT  
 GTAGCATATGAGGCACGGCGCTTATTTCTGACAAAATTTGTTGGTGCAAAAGGAATTC  
 ATTTATTTGACATCATTTTAACATCAGTGTTCAGGAGATTGGGGCTCAGACATATT  
 AGACAATATGTGATAGTGTCTACGTTACATGGGGAGCTCGGCATAATTCAGGAGCA  
 AGGGCAGCCCCAGGACAACCATACCTCCACATGGAACCACTTGGAAAACCTAAACT  
 CTACTGATCTCAAGGATGTTATTAAGGGGCTTATTCATTATGGACGAGATAACCA  
 GAATTTAGACATTTTACTTTTCCACGAAGTCTTGGAGTATATGTCTAGGATAGATAGA  
 GTGCTGAGTTTCCCTGGAGGTTCACTTCTATTAGCAGGACGCAGTGGTGTAGGTCGTC  
 GGACCATCACTTCTTTAGTCAGTCACATGCATGGAGCGGTCTGTTTCTCCAAGAT  
 TTCCAGAGGATATGAAGTGAAGCAGTTCAAAAATGATCTCAACATGTGCTGCAACTT  
 GCAGGAATTGAAGCACAACAGGTAGTTTTACTTCTTGAGGATTACCAAGTTTGTACATC  
 CTACATTTTTGGAGATGATCAATAGCCTTTGTCTTCAGGTGAAGTTCTGGACCTTA  
 TACTCTTGAAGAATTAGAGCCCTTGCTGTTACCCTTAAGGATCAAGCTTCAACAGAT  
 GGTTTTTTTGGACCACTTCAATTACTTCACATATAGAATTCAGCAAACTTGCATA  
 TTGTCTTGATAATGGATTCTGCAAATTCAACTTCATGATAAACTGTGAGAGTAATCC  
 AGCTTTCATGAAGAAATGCCAGGTGTTGTGGATGGAGGGTTGGTCCAATAGCAGTATG  
 AAGAAAATACCTGAAATGTTATTTCAGTGAAAACAGGTGGTGGAGAAAAATACAATGATA  
 AAAACGAAAAGAAGAAAAGAAAAAAATTCAGTTGATCTGATTTTCTGATTTTCAATCAT  
 TTTATTAATCCATGAATCTTGTAAAGCATATGGTGTACACCAAGCCGATACATGACC  
 TTTTACATGTGTATTCTGCCATTAGTAGTAGCAAGAAAAAGGAATTATTAAGAGAC  
 AAAGTCATTTGCAGGCTGGTGTATCTAACTAAATGAAGCTAAAGCTCTTGTGGATGA  
 ACTGAACAGAAAAGCTGGAGAACAAAGTGTGTTACTTAAAACGAAGCAAGATGAAGCA  
 GATGCTGCCCTTCAAATGATCAGTGTCAATGCAGGATGCTAGTGAGCAAAAAACAG  
 AACTTGAAAGACTGAAGCACAGAATAGCAGAAGAAGTTGTTAAAATGGAAGAAAGAAA  
 AAATAAAATGATGATGAATTAAGAAAGTACAACCTTTAGTCAATGAAGCTAAACTA

GCAGTTGGAAACATTAAGCCCGAATCACTTTTCAGAAATTCGCTCACTACGCATGCCAC  
CTGATGTAATTAGAGATATTCTTGAAGGAGTTTAAAGGTTGATGGGTATCTTTGATAC  
ATCTTGGGTGAGCATGAAAAGTTTCCTTGCAAAAAGAGGTGTAAGAGAAGACATAGCA  
ACCTTTGATGCCCGAAAATATTTCAAAGGAAATAAGAGAGAGTGTTGAAGAACTTCTTT  
TTAAAAATAAAGGCTCTTTTGATCCAAAGAAATGCTAAGCGTGCCAGTACTGCAGCTGC  
ACCTTTGGCTGCCCTGGGTGAAAGCCAATATTTCAGTATTCCTCATGTCTTGAACGAATT  
CATCTTTGGAACTGAACAGGCAGGATTAGAATCGAATCTGAAGAAAACCTGAAGACA  
GAAAAAGGAACTAGAGGAGCTTCTTAATTCGTGGTCAAAGGTATCAGAACTCAA  
AGAAAAATTTAGAGCAGGACTTCAGAAGCTGCCAAACTTGAGGCTGAAGTAAGCAAG  
GCACAAGAAACAATCAAAGCTGCAGAAGTCTTAATTAATCAGCTTGACAGAGAACATA  
AGAGATGGAATGCACAGGTTGTAGAGATAACAGAGGAATTAGCTACTCTTCTAAAAG  
AGCTCAACTTGCTGCTGCATTTATTACATATCTTTCTGCTGCTCCTGAATCTCTGAGA  
AAAACCTGTTTGAAGAATGGACCAAGTCAGCTGGTCTTGAGAAATTTGATCTGAGGA  
GATTTCTTTGTACTGAAAGTGAGCAGTTAATTTGGAAAAGTGAAGGCCCTACCATCAGA  
TGACCTTTCCATAGAAAATGCTCTTGTAATATTACAGAGTCGAGTGTGCCATTCTCTT  
ATAGATCCTTCTTCCCAAGCTACAGAGTGGTTAAAAACACATTTGAAAGACTCAGCTT  
TAGAAGTTATCAATCAGCAGGATAGTAACTTTATCACAGCTCTTGAATTAGCAGTACG  
TTTTGGGAAAACCTTATTATACAAGAGATGGATGGTGTAGAACCTGTTCTTTATCCA  
TTATTGAGACGAGATCTGGTTGCTCAAGGACCACGTTATGTGGTACAAATAGGTGACA  
AAATTATTGACTACAATGAAGAATTCCGCCCTCTTTTGTCAACAAGAAACCCAAATCC  
TTTTATTCCACCGGATGCAGCTTCCATTGTTACTGAGGTTAACTTTACTACAACAAGA  
AGTGGATTACGAGGCGAGCTTTTAGCTTTAACCATTCAGCATGAGAAACCTGATTAG  
AAGAACAGAAAACAAAACCTATTACAACAGGAAGAAGATAAGAAAATACAGCTAGCCAA  
GCTCGAAGAATCTCTCTAGAGACACTTGCCACATCTCAAGGCAATATTTTGGAAAAT  
AAGGATTGATTGAGTCTTTGAATCAGACAAAAGCAAGCAGTGCACTTATTCAAGAGT  
CACTTAAAGAATCTTACAACTCCAAATTTCCCTTGATCAAGAACGGGATGCCTATCT  
CCCCCTGGCTGAGAGTGCCAGCAAGATGTACTTCATTATTTCTGATTTGTCCAAAAT  
AATAACATGTACCGTTTGTAGTTGGCTGCTTTTCTCCGACTTTTCCAACGAGCTCTAC  
AAAACAAACAGGATTCTGAAAATACAGAACAGAGAATCCAGTCACTTATCAGCTCAT  
ACAACATATGGTATATGAATATATATGTCGTTGTCTATTTAAGGCTGATCAGTTGATG  
TTCGCTTTGCATTTTGTTCGAGGCATGCATCCTGAACCTTTTCAAGAAAATGAATGGG  
ATACGTTTACAGGTGTGGTTGTTGGAGACATGTTACGGAAAGCTGACTCAACATAA  
AATACGTGATCAGCTTCCGCTCTGGATAGATCAGGAACGAAGCTGGGCCGTGGCAACA  
TTAAAGATTGCTCTCCCGAGTCTTTATCAGACCTCTGCTTTGAAGATGCAGCTCTGT  
GGCGTACTTATTATAATAATTCAATGTGTGAGCAAGAGTTCCATCTATCCTTGCAAA  
GAAAGTTTCCTTATTTTCAGCAGATTCTTGTAAGTACAGGCGCTAAGACCCGGACAGATTG  
CAAAGTGCCATGGCTCTTTTGCATGTAAACTCTGGGACTGAAAGAGGTGTCCCCAC  
TGCCCTCTAAATCTCAAACGTTTATACAAAGAGACACTGGAATTTGAACCATCTTGAT  
AATTATTTCTCCGGGTGCTGATCCTTCTCAGGAACCTCAAGAACTAGCTAATGCTGAA  
AGAAGCGGAGAGTGTATCACCAGGTTGCCATGGGTCAAGGTCAAGCTGATTTAGCAA  
TTCAAATGCTAAAAGAATGTGCCCGCAATGGAGACTGGCTCTGTTTGAAGAACTTACA  
TCTTGTGGTATCTTGGCTGCCAGTCTTGGAAAAGGAATTGAATACTCTTCAACCTAAA  
GATACCTTTGCTCTTTGGCTCACTGCAGAAAGTTCATCCCACTTTACTCCTATTTTAC  
TACAGTCAAGTCTGAAGATAACATATGAGTCACCTCCAGGTTTAGAGAAGAATTTAAT  
GCGTACTTATGAGTCTTGGACTCCTGAGCAAATTAGCAAAAAGATAATACACATCGA  
GCTCATGCTCTCTCAGTCTTGCATGGTTTCATGCTGCATGTCAAGAAAGAAGAACT  
ATATTCCTCAGGGTTGGACAAAGTTTTATGAATTTCTTTATCAGATCTTCGGGCTGG  
GTACAACATTATAGACAGCTTTTGTATGGTGCCAAAGATGTACAATGGGAATTTGTA  
CATGGTTTACTTGAAAATGCTATTTATGGAGGACGTATAGACAACCTATTTGACCTTA  
GAGTTCTTCAGTCATACCTGAAGCAGTTTTTTAATTCCTCAGTTATTGATGTATTCAA  
CCAAAGGAACAAGAAAAGCATTTTTCATATTCGATCTCTACCACAATCCTGCAGC  
ATTTTGAGACTATCGTGTCTGATTGAGAAAATTCAGAGGACGACAAACCTAGTTTCT  
TTGGTCTGCCTGCCAATATCGCTCGCTCATCTCAGCGCATGATCAGTTCTCAGGTTAT  
TTCACAGTTGAGGATTTTGGGCAGATCCATAACAGCTGGTTCCAAATTTGATAGAGAA  
ATCTGGTCTAATGAACCTTTCTCCTGTCTCAATCTCTGGAAGAACTAAACAGAAAT  
CAAACCTAATACATCAGAAAGTGCCTCCTCCTAACGATCGACAAGGATCTCCAATACT  
GTCATTATCATATTCTTGAACAATTTAATGCTATTGTTTGTAGTACAAAGTGTCCACCAG  
TCTCTTGCTGCTCTCAGCAAAGTCATCAGAGGAACCTACTTTACTGAGTTCAGAAGTAC  
AAAAATGGCAAGTGCTTTATTAAACCAAAGTGCTCTCTCGCATGGCAGAGCAAGTG  
GGAAGGCCAGAAAGATCCCTTACAATACCTGAGAGGCTCTTGTGCCCCGTGCCCTGCA  
ATACAGAACTGGGTAGATAAAGCTGAAAAACAGGCTCTTCTCTGAAACACTTGACC

	TATCAGAACTTTTCCATCCAGACACATTTCTTAATGCTCTTCGCCAGGAACTGCAAG GGCAGTGGGTCGTTCTGTGGATAGCCTTAAATTTGTAGCCTCATGGAAGGTCGACTG CAAGAAGCAAAGCTACAAATTAAGATCAGTGGCTTGTACTAGAAGGATGTAGTTTTG ATGGAAATCAACTTTCTGAAAATCAGCTTGATTCTCCAGCGTGTATCAGTGCTCCC TTGTTTTATGGGCTGGATTCCACAGGATGCATGTGGTCCATATTCTCCGGATGAGTGC ATCTCTTTGCCTGTTTACACAAGTGCTGAAAGGGATCGTGTGGTTACCAATATTGATG TTCCATGTGGGGCAACCAAGACCAGTGGATTCACTGTGGAGCAGCTCTATTCTCTAAA AAATCAGTAGAATCTAATGACAACAAAAGCCATCTTCACAAAAGGGAACATTGATTCT TTAAGCTTTAAATCAAACATGTGGTCACTCTACATTTGAAATGTAGTTCAAATATT AACATATAGTTATGTTGTTGATGTCACTGAAATTTAATGTGTAAGCAGCAGCTGTG CATCTTTTAAAGTAATAAATTAATGGAGTTATTGTTAAACAGAGTATTCTTTTGACA ACATTAAATATTTCTGTGAGAAAGTCACTTTTCCAGTGGCTCAAAAATTTGTTTAG GTCAGAGATTTTAAGTGGTATATTAACCAATAATAAATATTTTGCTGTC		
	ORF Start: ATG at 61		ORF Stop: TAG at 13000
	SEQ ID NO: 312	4313 aa	MW at 493435.2kD
NOV29a, CG154509-01 Protein Sequence	MANGTADVRKLFIFTTQNYFGLMSELWDQPLLCNLEINNFLDDGNQMLLRVQRSDA GISFSNTIEFGDTKDKVLVFFKLRPEVITDENLHDNILVSSMLSEPISSLYQAVRQVF APMLLKDQWSRNFDPKLQNLLESEAGLGIVLRRSDTNLTCLKFKEDDTRGILTPSD EFQFWIEQAHRGNKQISKERANYFKELFETIAREFYNLDSLLEVVLDVETQDVVD DVWRQTEHDHYPESRMLHLDDIIGGSFGRFVQKLGTLNLWEDPYLVKESLKAGISI CEQWVIVCNHLTGQVWQRYVPHPWKNEKYFPETLDKLGKRLBEEVLAIRTIHEKFLYFL PASEEKIICLTRVFEPTGLNPVQYNPYTEPLWKAASVQYEKIIPAEQKIAGKLKNY ISEIQDSPQQLLQAFLYKELVKRPTISKELMLERETLLARLVDISKDFRLDFENRCR GIPGDASGPLSGKNLSEVVNSIVWVRQLELKVDDTIKTAELSLDLPGRFCFHQSAKD LLDQLKLYEQEQFDDWSRDIQSGLSDSRSGLCIEASSRIMELDSNDGLLVHYSRDLV ILLREVRQLSALGFVPAKIQQVANIAQKFKQAIILKQVAHFYNSIDQQMIQSQRPM MLQSALAFEQIKNKSKAGSGGKSQITWDNPKLEGYIQKLQNAERLATENRKLKRW TTFCEKVVLNMNIDLLRQQQRWKDGLQELRTGLATVBAQGFQASDMHAWKHWNHOLY KALEHQYQMGLEALNENLPEINIDLTQKQRLQFRPPFEEIRAKYREMKRFIGIPNQ FKGVGEARSINSIFSIMIDRNASGFLTIFSKAEHLFRRLSAVLHQHKEWIVIGQVDM ALVEKHLFTVHDWEKNFKALKIKGKEVERLPSAVKVDCLNINCNPVKTVIDDLIQKLF DLLVLSLKKSIIQAHLHEIDTFVTEAMEVLTIMPQSVEEIGDANLQYSKLQERKPEILP LFQEAEDKNRLLRTVAGGGLLETISNLKAKWKDFELMMESHQMIKDQIEVMKGNVKS LQIYYQELEKFKARWDQLKPGDDVIETGQHNTLDKSAKLIKKEKIEFDDLEBTRKKLV DDCHHFRLEEPNFSLASSISKDIESCAQIWAFFYEEFQQGFQEMANEDWITFRTKTYLF EEFLMNWHDRLRKVEEHSVMTVKLQSEVDKYKIVIPILKYVRGEHLSPDHWDLDFRLL GLPRGTSLEKLLFGDLLRVADTIVAKAADLKDLSRAQGEVTIREALRELDLWGVGAV FTLIDYEDSQSRMTKLKDWKDIVNQVDNRCLLQSLKDSPPYKGFEDKVS IWERKLA ELDEYLQNLNHIQRKWVYLEPIFGRGALPKEQTRFNVRVEDFRSIMTDIKKDNRVTTL TTHAGIRNSLLTILDQLQRCQSLNEFLKRSAPFRFYFIGDDLLEILGQSTNPSV IQSHLKKLFAGINSVCDEKSKHITAMKSLEGEVVPFKNKVPLSNNVETWLNDALEM KKTLEQLLKECVTTGRSSQGAVDPSLFPSQILCLAEQIKFTEDVENAIKDHSLHQIET QLVNKLEQYTNIDTSSDPGNTESGILELKLKALILDIHNIDVVKQLNQIQVHTTED WAWKKQLRPFYMKSDHTCCVQMVDSFQYTYEYQGNASKLVYPLTDKCYLTLTQAMKM GLGCPYGPAGTGKTESVKALGGLGRQVLVFNCDGIDVKSMBGRIFVGLVKCGAWGC FDEFNRLEESVLSAVSMQIQTIQDALKNHRTVCELLGKEVEVNSNSGIFITMNPAGKG YGGRQKLPDNLKQLFRPVAMSHPDNELIAEVILYSEGFKDAKVLRSKLVAIFNLSREL LTPQQHYDWGLRALKTVLRGSGNLLRQLNKSQTTQANESHIVVQALRLNMTSKFTFT DCTRFDALIKDVFPGLIELKEVEYDELSAALKQVFEEANYEIIIPNKKALEBLYEQLCQ RMGVVIVGSPGAGKSTLWRMLRAALCKTGKVVQYTMNPKAMPYQLLGHIDMDTREW SDGVLTSARQVVREPQDVSSWIIDGDIDPEWIESLNSVLDDNRLLTMPSGERIQFG PNVNFVFETHDLSCASPATISRMGMIFLSDEETDLNSLIKSWLRNQPAEYRNNLENWI GDYFEKALQWVLKQNDYVVTSLVGTVMNGLSHLHGRDHDEFIINLIRGLGGLNLMK SRLEFTKEVFHWARESPDFHKPMDTYDSTRGLATYVLKPEDLTADDFSNGLTLP VIQTPDMQRGLDYFKPWLSSDTKQPFILVGPPEGCGKGMLLRYAFSQRSTQIATVHCS AQTTSRHLLQKLSQTCMVISNTGRVYRPKDCERLVLYLKDINLPKLDKWTSTLVAF LQQVLTYYQGFYDENLEWVGLENIQIVASMSAGRLGRHKL'TRFTSIVRLCSIDYPER EQLQTIYGAYLEPVLHKNLKNHSIWGSSSKIYLLAGSMVQVYEQVDMHQVRAKFTVDD YSHYFFTPCILTQWVLGLFRYDLEGGSSNHPLDYVLEIVAYEARRLFRDKIVGAKELH		

	LFDIILTSVFQGDWGSIDLDNMSDSFYVTWGARHNSGARAAPGQPLPPHGKPLGKLS TDLKDVIKKGLIHYGRDNQNLIDILFHEVLEYMSRIDRVLSFPGGSLLAGRSGVGRR TITSLVSHMHGAVLFSFKISRGYELKQFKNDLKHVLQLAGIEAQQVLLLEDYQFVHP TFLEMINSLLSSGEVPGLYTLEELPLLLPLKDQASQDGGFFGPVFNFTYRIQQNLHI VLIMDSANSNFMNCESNPALHKKCQVLWMEGWSNSSMKKIPEMLFSETGGGEKYNDK KRKEEKKNSVDPDFLKSFLLIHESCKAYGATPSRYMTFLHVYSAISSKKKELLKRQ SHLQAGVSKLNEAKALVDELNRKAGEQSVLLKTKQDEADAALQMITVSMQDASEQKTE LERLKHRIAEEVVKIEERKNKIDDELKEVQPLVNEAKLAVGNIKPESLSEIRSLRMP DVIRDILEGVLRLMGIFDTSWVSMKSFLAKRGVREDIATFDARNISKEIRESVEELLF KNKGSFDPKNAKRASTAAAPLAWVKANIQYSHVLERIHPLETEQAGLESNLKKTEDR KRKLELLNSVGQKSELKEKFSRTSEAAKLEAEVSKAQETIKAAEVLINQLDREHK RWNAQVVEITEELATLPKRAQLAAAFITYLSAAPESLRKTCLEEWTKSAGLEKFDLRR FLCTESEQLIWKSEGLPSDDLSENALVILQSRVCPFLIDPSSQATEWLKTHLKDSRL EVINQQDSNFITALELAVRFGKTLIIQEMDGVPEVLYPLLRDLVAQGPRYVQIGDK IIDYNEEFRLFLSTRNPNPFIIPDAASIVTEVNFTTTRSLRGLQQLALTIQHEKPDLE EQKTKLLQOEEDKKIQLAKLEESLLETATLSQGNILENKDLIESLNQTKASSALIQES LKESYKLQISLDQERDAYLPLAESASKMYFIISDLSKINNMYRFSALFLRLFORALQ NKQDSENTEQRIQSLISSLOHMYEYICRCLFKADQLMFALHVRGMHPELFQENEWD TFTGVVVGDMRLKADSQQKIRDLQPSWIDQERSWAVATLKIALPSLYQTLCFEDAALW RTYYNSMCEQEFPSSILAKKVSILFQQILVVQALRPDLQSAMALFACKTLGLKEVSP PLNLKRLYKETLEIEPILIIISPGADPSQELQELANAERSGECYHQVAMQGGQADLAI QMLKECARNGDWLCNLHLVSVLWLVLEKEKELNTLPKDTFLRLWLTAEVHPNFTPI QSSLKITYESPPGLEKNLMRTYESWTPEQISKDNTHRAHALFSLAWFHAACQERRNY IPQGWTKFYEFSLSDLRAGYNIIDRLFDGAKDVQWEFVHGLLENAIYGRIDNYFDLR VLQSYLKQFFNSSVIDVFNQRNKKSIFFPYSVSLPQSCSILDYRAVIEKIPEDDKPSFF GLPANIRSSQRMISQVISQLRILGRSITAGSKFDREIWSNELSPVLNLWKKLNQNS NLIHQKVPVPPNDRQGSPIILSFIIIEQFNAILRVQSVHQSALASKVIRGTTLSSSEVQ KLASALLNQKCPALWQSKWEGPEDPLQYLRGLVARALAIQNWVDKAEQALLSETLDL SELFHPDTFLNALRQETARAVGRSVDSLKFVASWKGRLQEAQLQIKISGLLLEGCSFD GNQLSENQLDSPSVSSVLPFCFMGWIPODACGPYSPDECISLPVYTSERDRVVTNIDV PCGGNQDQWIOCGAALFLKNQ
--	---

Further analysis of the NOV29a protein yielded the following properties shown in Table 29B.

Table 29B. Protein Sequence Properties NOV29a	
PSort analysis:	0.6000 probability located in nucleus; 0.3600 probability located in mitochondrial matrix space; 0.3249 probability located in microbody (peroxisome); 0.1000 probability located in lysosome (lumen)
SignalP analysis:	No Known Signal Sequence Predicted

5 A search of the NOV29a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 29C.

Table 29C. Geneseq Results for NOV29a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV29a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
ABB70206	Drosophila melanogaster	55..2085	708/2074 (34%)	0.0

	polypeptide SEQ ID NO 37410 - <i>Drosophila melanogaster</i> , 2055 aa. [WO200171042-A2, 27-SEP-2001]	20..2015	1159/2074 (55%)	
ABB60101	<i>Drosophila melanogaster</i> polypeptide SEQ ID NO 7095 - <i>Drosophila melanogaster</i> , 4472 aa. [WO200171042-A2, 27-SEP-2001]	896..4311 1081..4471	959/3550 (27%) 1674/3550 (47%)	0.0
AAB93815	Human protein sequence SEQ ID NO:13606 - <i>Homo sapiens</i> , 553 aa. [EP1074617-A2, 07-FEB-2001]	3761..4313 1..553	551/553 (99%) 552/553 (99%)	0.0
AAM79140	Human protein SEQ ID NO 1802 - <i>Homo sapiens</i> , 2166 aa. [WO200157190-A2, 09-AUG-2001]	2193..4299 14..2151	612/2209 (27%) 1078/2209 (48%)	0.0
AAM80124	Human protein SEQ ID NO 3770 - <i>Homo sapiens</i> , 2088 aa. [WO200157190-A2, 09-AUG-2001]	2263..4299 9..2073	596/2135 (27%) 1048/2135 (48%)	0.0

In a BLAST search of public sequence databases, the NOV29a protein was found to have homology to the proteins shown in the BLASTP data in Table 29D.

Table 29D. Public BLASTP Results for NOV29a				
Protein Accession Number	Protein/Organism/Length	NOV29a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q9JJ79	Cytoplasmic dynein heavy chain - <i>Rattus norvegicus</i> (Rat), 4306 aa.	1..4313 1..4306	4004/4313 (92%) 4175/4313 (95%)	0.0
Q27802	Dynein heavy chain isotype 1B (EC 3.6.1.3) - <i>Tripleneustes gratilla</i> (Hawaiian sea urchin), 4318 aa.	7..4313 5..4318	2677/4338 (61%) 3354/4338 (76%)	0.0
Q19542	F18C12.1 protein - <i>Caenorhabditis elegans</i> , 4131 aa.	1..4311 1..4131	1719/4328 (39%) 2570/4328 (58%)	0.0
BAC02706	KIAA1997 protein - <i>Homo sapiens</i> (Human), 1194 aa (fragment).	3120..4313 1..1194	1192/1194 (99%) 1193/1194 (99%)	0.0
Q9SMH5	Cytoplasmic dynein heavy chain	39..3064	1249/3133 (39%)	0.0

	1b - Chlamydomonas reinhardtii, 3074 aa (fragment).	39..3074	1833/3133 (57%)	
--	--	----------	-----------------	--

Pfam analysis predicts that the NOV29a protein contains the domains shown in the Table 29E.

Table 29E. Domain Analysis of NOV29a			
Pfam Domain	NOV29a Match Region	Identities/ Similarities for the Matched Region	Expect Value
PRK	1976..2002	9/28 (32%) 20/28 (71%)	0.69
DUF164	3099..3307	52/239 (22%) 112/239 (47%)	0.15
Dynein_heavy	3613..4311	218/790 (28%) 513/790 (65%)	9.9e-129

### Example 30.

The NOV30 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 30A.

5

Table 30A. NOV30 Sequence Analysis			
	SEQ ID NO: 313	4292 bp	
NOV30a, CG155595-01 DNA Sequence	GTCAGGAGCTGCAGGATCTGGCTCGAGTCCCCTGCAGGGGCCAGAGCAGTCTCCCT CGGCATGGGGCTGGAGGCTCAGAGGCTGCCAGGGGCTGAGGAGGCCCAAGTGC GGTT GCCCTGCGAGTTCGACCACTGCTGCCCAAGGAGCTGCTGCACGGGCATCAGAGCTGCC TGCAGGTGGAGCCAGGGCTTGGCCGCGTCACTCTGGGCCGTGACCGACACTTTGGCTT CCACGTGGTGCTGGCCGAGGATGCGGGGCAGGAGCCGTGTACCAGGCTGCGTTGAG CCCCCTCTTGAAGCCTTCTTCGAGGGCTTCAATGCCACTGTCTTTGCCATATGGTCAGA CGGGCTCAGGGAAGACATACACCATGGGGGAGGCCAGTGTGGCCTCCCTCCTTGAGGA TGAGCAGGGCATTGTCCCAGGGCCATGGCCGAGGCCCTCAAGCTCATCGATGAGAAC GACCTGCTTGACTGTCTGGTACATGTGTCTACCTGGAAGTGTACAAGGAGGAGTTCC GAGACCTGCTCGAGGTGGGCACTGCCAGCCGTGACATCCAGCTCCGGGAAGATGAGCG CGGGAATGTTGTCTGTGCGGGGTGAAGGAGGTGACGTGGAGGGCCTGGATGAGGTG CTGAGCCTCCTGGAGATGGGCAACGCGGCGCGGCACACGGGAGCCACGCACCTCAACC ACCTGTCTAGCCGCTCACACACGGTCTTACCGTGACCTGGAGCAGCGGGGGCGCGC CCCCAGCCGCTACCCGCCCCCGCCCCGGGCCAGCTGCTCGTCTCCAAGTTCCACTTC GTGGACCTGGCGGGCTCAGAGAGGGTGCTCAAGACGGGCAGCACCGGCGAGCGGCTCA AGGAGAGCATCCAGATCAACAGCAGCCTCCTGGCGCTGGGCAACGTATCAGCGCCCT GGGGACCTCAGCGCCGGGGCAGCCACATACCCTACCGGCACTCCAAGATCACCCGG ATCTCAAAGACTCGCTGGCGGGAACGCAAGACGGTGATGATCGCTGCGTCAGCC CTTCTCTCTCCGACTTCGACGAGACCTCAACACCTCAACTACGCCAGCCGCGCCCA GAACATCCGCAACCGCGCCACGGTCAACTGGCGGCCGAGGCCGAGCGGCCACCCGAA GAGACGGCGAGCGGCGCGGGGTCCGCCACGGCACCGCTCCGAGACCCGATCATCC ACCGCGGCCGCGCGCCCCAGGCCACCGCCTCCGCGGCGGCGCCATGCGCCT GGGCGCCGAGTGC GCGCTACCGGGCTGCACCGACGCGCCTACAGCCTCTTGCGC GAGCTGCAGGCCGAGCCCGGGCTGCCCCGCGCGCCGCCGCAAGGTGCGCGACTGGC TGTGCGCCGTGAGGGGCGAGCGCAGCGCCCTGAGCTCCGCCTCCGGGCCGATAGCGG CATCGAGAGCGCTCCGTGAGGACAGGCGGCGCAGGGGCCGCGGGCGAAAGGTG		

	GCCGAGGGACAGGAGGATGAGGGGGCGCAGCAGCTGCTGACCTGCAGAACCAGGTGG CGCGGCTGGAGGAGGAGAACCAGACTTTCTGGCTGCGCTGGAGGACGCCATGGAGCA GTACAAACTGCAGAGCGACCGGCTGCGTGAGCAGCAGGAGAGATGGTGGAACTGCGG CTGCGGTTAGAGCTGCTGCGGCCAGGCTGGGGGGGCGCGGCTCCTGAATGGCCTGC CTCCCGGGTCTTTGTGCCTCGACCTCATACAGCCCCCTGGGGGGTGGCCACGCCCA TGTGCTGGGCATGGTGCCGCTGCCTGCCTCCCTGGAGATGAAGTTGGCTCTGAGCAG AGGGGAGAGGTGACAAATGGCAGGGAGGCTGGAGCTGAGTTGCTGACTGAGGTGAACA GGCTGGGAAGTGGCTCTTCAGCTGCTTCAGAGGAGGAAGAGGAGGAGGAGGCCGCC CAGGCGGACCTTACACCTGCGCAGTTGGGGCAGCAACCTTGACAGGCTGCCTGTTGCA GCAGTTGGTGGGAGCAAGGCCGAGTTCAGGCCCGCCAGGTCCCCCTGCCACAGCCT CAGAGTGGCGGCTGGCCAGGCCAGCAGAAGATCCGGGAGCTGGCTATCAACATCCG CATGAAGGAGGAGCTTATTGGCGAGCTGGTCCGCACAGGAAAGGCAGCTCAGGCCCTG AACCGCCAGCACAGCCAGCGTATCCGGGAGCTGGAGCAGGAGGCAGAGCAGGTGCGGG CCGAGCTGAGTGAAGGCCAGAGGCAGCTGCGGGAGCTCGAGGCAAGGAGTCCAGGA TGCTGGCGAGCGGTCTCGGCTCCAGAGTTCGCGCAGGAGGGTGCCTGCGGCCAGAGC CAGGTGCAGGTGCTGAAGGAGAAGAAGCAGGCTACGGAGCGGCTGGTGTCACTGTGCG CCCAGAGTGAGAAGCGACTGCAGGAGCTCGAGCGGAACGTGCAGCTCATGCCGCAGCA GCAGGGACAGCTGCAGAGGCGGCTTCGCGAGGAGACGGAGCAGAAGCGGCGCTGGAG GCAGAAATGAGCAAGCGGCAGCACCGCGTCAAGGAGCTGGAGCTGAAGCATGAGCAAC AGCAGAAGATCCTGAAGATTAAGACGGAAGAGATCGCGGCATTCCAGAGGAAGAGGCG CAGTGGCAGCAACGGCTCTGTGGTTCAGCCTGGAACAGCAGCAGGTGGGGCCAGGCTGT GTCCGCACCCAGGGCTCCCTTGGGGCTGGCTGGTGGGTGCACCTTTCTCCCCAGTGA ACCTCGAGTGGCGGCTGACACAGCCAGAGAAGATTGAGGAGCAGAAGAAGTGGCTGGA CCAGGAGATGGAGAAGGTGCTACAGCAGCGCGGGCGCTGGAGGAGCTGGGGGAGGAG CTCCACAAGCGGGAGGCCATCCTGGCCAAGAAGGAGGCCCTGATGCAGGAGAAGACGG GGCTGGAGAGCAAGCGCCTGAGATCCAGCCAGGCCCTCAACGAGGACATCGTGCGAGT GTCCAGCCGGCTGGAGCACCTGGAGAAGGAGCTGTCCGAGAAGACGCGGCAGCTGCGG CAGGGCAGCGCCAGAGCCAGCAGCAGATCCGCGGGGAGATCCACAGCCTGCGCCAGG AGAAGGACTCGTGTCTCAAGCAGCGCCTGGAGATCGACGGCAAGCTGAGGCAGGGGAG TCTGCTGTCCCCGAGGAGGAGCGGACGCTGTTCCAGTTGGATGAGGCCATCGAGGCC CTGGATGCTGCCATTGAGTATAAAGATGAGGCCATCACATGCCGCCAGCGGCTGCTTC GGGCTCAGCCTCGTTGCTGTCCAGTGCGAGATGAACCTCATGGCCAAGCTCAGCTA CCTCTCATCCTCAGAGACCAGAGCCCTCCTCTGCAAGTATTTGACAAGGTGGTGAGC CTCCGAGAGGAGCAGCACACAGCAGCAGATTGCCTTCTCGGAAGTGGAGATGCAGCTGG AGGAGCAGCAGAGGCTGGTGTACTGGCTGGAGGTGGCCCTGGAGCGGCAGCGCCTGGA GATGGACCGCCAGCTGACCTGCGAGCAGAAGGAGCAGAGCAGAACATGCAGCTGCTC CTGCAGCAGAGTCGAGACCACCTCGGTGAAGGGTTAGCAGACAGCAGGAGGCAGTATG AGGCCCGGATTCAAGCTCTGGAGAAGGAACTGGGCCGTTACATGTGGATAAACACAGGA ACTGAAACAGAAGCTCGGCGGTGTGAACGCTGTAGGCCACAGCAGGGGTGGGGAGAAG AGGAGCCTGTGCTCGGAGGGCAGACAGGCTCCTGGAATGAAGATGAGCTCCACCTGG CACCCGAGCTTCTTGGCTGTCCCCCTCACTGAGGGGGCCCCCGCACCCGGGAGGA GACGCGGGAAGTGGTCCACGCTCCGTTACCTTGACCTGGAACGCTCGAGCCTGTGT GGGACTCTTCAACAACACCAATATCAGGACCAGGATCAGAGGACCTCGAGGAACAC ATGCACAAGGATTATTCCATACCACTTGTAATTAAACATTATTAAGGAGACAGGCAGC TTCTCACTTAACAAGATCACAAAGATCACAGGCTGTGATAACACCAGTGTCTATTTC TGAAATGTGGTACCTTTGTTCTTCTGAAGTTGTCAAGTTTATCCTCTAGACCATCCA CAGCTGACACAGAATGGCTTCTAGGCAACCCCCGCTTTAGTGATCTCTTTGAAGGGGA AAGCAATTCTGGTTGAAAAGATTCTTCAACTTTGGTCACTTCTAAAAGCATCAAA		
	ORF Start: ATG at 63	ORF Stop: TAA at 4035	
	SEQ ID NO: 314	1324 aa	MW at 148066.3kD
NOV30a, CG155595-01 Protein Sequence	MGLEAQRPLPGABEAPVRVALRVRPLLPKELLHGHQSLQVEPGLGRVTLGRDRHFGFH VVLAEADAGQEAVYQACVQPLLEAFFEGFNATVFAYGQTGSGKTYTMGEASVASLLEDE QGIVPRAMAEAFKLIDENLLDCLVHVSYLEVYKEEFRDLLEVGTASRDIQLREDERG NVVLGCVKEVDVEGLDEVLSLLEMGNARHTGATHNLSSRSHTVFTVTLEQRGRAP SRLPRPAPGQLLVSKFHFVDLAGSERVLKGTSTGERLKESIQINSSLLALGNVISALG DPQRRGSHI PYRDSKITRILKDSLGGNAKTMIACVSPSSSDFDETNLNLNYASRAQN IRNRATVNRPEAERPPEETASGARGPPRHRSETRIIHRGRRAPGPATASAAAAMRLG AECARYRACTDAAYSLLRELQAEPLPGAAARKVRDWLCAVEGERSALSASGPDSGI ESASVEDQAAQGAGGRKVAEGQEDEGAQQLLTQNQVARLEENRDFLALEDAMEQY		

KLQSDRLREQQEEMVELRLRLLELVRPGWGGPRLLNGLPPGSFVPRPHTAPLGGAHAHV LGMVPPACLPGEVSGEQRGEVTNGREAGAELLTEVNRLGSGSSAASEEEEEEEPPR RTLHLRSWGSNLDRLPVAAVGGSKARVQARQVPPATASEWRLAQAAQKIRELAINIRM KEELIGELVRTGKAAQALNRQHSQRIRELEQEAEQVRAELSEGQRQLRELEGKELQDA GERSRLQEFRRRVAAAQSQVQLKEKKQATERLVSLSAQSEKRLQELERNVQIMRQQQ GQLQRRRLREBTEQKRRLEAEMSKRQHRVKELELKHEQQQKILKIKTEEIAAFQKRRS GSNGSVVSLQQQVGPVCVTRTQGSPPGWLVGAPFSPVNLWRLTQPEKIEEQKKWLDQ EMEKVLQQRRALEELGEELHKREAILAKKEALMQEKTGLESKRRLRSSQALNEDIVRVS SRLEHLEKELSEKSGQLRQGSAAQSQQIRGEIDSLRQEKDSLKQRLKIDGKLRQGS LSPEEERTLFLDEAIEALDAAIEYKNEAITCRQVLRASASLLSQCEMNLMAKLSYL SSSETRALCKYFDKVVTLREEQHQQQIAFSEMQLEEQQRLVYVLEVALERQRLM DRQLTLQQKEHEQNMQLLLQQSRDHLGEGGLADSRQYEARIQALEKELGRYMWINQEL KQKLGGVNAVGHSGGKRSKSLCSEGRQAPGNEDELHLAPPELLWLSPLTEGAPRTREET RDLVHAPLPLTWKRSSLCGDSSTTPISGPGSEDLPEPHAQGLFHTTCN
---

Further analysis of the NOV30a protein yielded the following properties shown in Table 30B.

Table 30B. Protein Sequence Properties NOV30a	
PSort analysis:	0.8800 probability located in nucleus; 0.3000 probability located in microbody (peroxisome); 0.1000 probability located in mitochondrial matrix space; 0.1000 probability located in lysosome (lumen)
SignalP analysis:	No Known Signal Sequence Predicted

5. A search of the NOV30a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 30C.

Table 30C. Geneseq Results for NOV30a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV30a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAU86160	Human PRO539 polypeptide - Homo sapiens, 830 aa. [WO200153486-A1, 26-JUL-2001]	519..1301 1..777	734/811 (90%) 737/811 (90%)	0.0
AAU96730	PRO539, a Costal-2 homologue - Homo sapiens, 830 aa. [WO200036102-A2, 22-JUN-2000]	519..1301 1..777	734/811 (90%) 737/811 (90%)	0.0
ABB81633	Human kinesin motor protein HsKif7 fragment SEQ ID NO:2 - Homo sapiens, 342 aa. [US6395527-B1, 28-MAY-2002]	11..354 1..342	341/344 (99%) 342/344 (99%)	0.0
ABB81634	Human kinesin motor protein HsKif7 fragment SEQ ID NO:4 - Homo sapiens, 337 aa.	12..350 1..337	336/339 (99%) 337/339 (99%)	0.0

	[US6395527-B1, 28-MAY-2002]			
ABB80078	Human kinesin motor protein (HsKrp5) amino acid sequence - Homo sapiens, 1279 aa. [US6379941-B1, 30-APR-2002]	676..1222 593..1102	259/548 (47%) 386/548 (70%)	e-131

In a BLAST search of public sequence databases, the NOV30a protein was found to have homology to the proteins shown in the BLASTP data in Table 30D.

Table 30D. Public BLASTP Results for NOV30a				
Protein Accession Number	Protein/Organism/Length	NOV30a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q95LL1	Hypothetical 98.5 kDa protein - Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey), 865 aa (fragment).	12..825 2..865	359/877 (40%) 527/877 (59%)	e-166
Q9UF54	Hypothetical 96.7 kDa protein - Homo sapiens (Human), 833 aa (fragment).	676..1222 147..656	256/548 (46%) 384/548 (69%)	e-129
Q9QXL2	Kif21a - Mus musculus (Mouse), 1573 aa.	8..356 2..378	178/377 (47%) 236/377 (62%)	2e-88
Q9CTY0	Kinesin family member 21A - Mus musculus (Mouse), 647 aa (fragment).	5..356 82..461	178/380 (46%) 236/380 (61%)	1e-87
Q9NXU4	CDNA FLJ20052 fis, clone COL00777 - Homo sapiens (Human), 576 aa (fragment).	8..356 2..378	175/377 (46%) 237/377 (62%)	8e-87

PFam analysis predicts that the NOV30a protein contains the domains shown in the Table 30E.

Table 30E. Domain Analysis of NOV30a			
Pfam Domain	NOV30a Match Region	Identities/ Similarities for the Matched Region	Expect Value
kinesin	21..364	168/404 (42%) 260/404 (64%)	1.3e-125
DUF164	681..913	55/251 (22%) 132/251 (53%)	0.015

**Example 31.**

The NOV31 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 31A.

Table 31A. NOV31 Sequence Analysis			
	SEQ ID NO: 315	5460 bp	
NOV31a, CG155962-01 DNA Sequence	ATGTCGGGAGCCTCAGTGAAGGTGGCTGTCCGGGTAAGGCCCTTCAATTCTCGAGAGA CCAGCAAGGAATCCAAATGCATCATTAGATGCAAGGCAACTCGACCAGTATTATTAA CCCAAAGAATCCAAAGGAAGCTCCAAAGTCCTTCAGCTTCGACTATTCTACTGGTCT CATACCTCACCCGAAGATCCCTGTTTTCATCTCAAAACCGTGTGTACAATGACATTG GCAAGGAAATGCTCTTACACGCCCTTGGAGGATATAATGTCTGTATTTTGCCTATGG GCAGACTGGTGTGGAATCTTATACAATGATGGGTAAACAAGAAGAAAGCCAGGCT GGCATCATTCCACAGTTATGTGAAGAACTTTTGGAGAAATCAATGACAACCTGTAATG AAGAAATGTCTTACTCTGTAGAGGTGAGTTACATGGAAATTTACTGTGAAAGAGTACG AGATTTGCTGAATCCAAAAACAAGGGTAATTTGCGTGTGCGTGAACACCCACTTCTT GGACCCATGTGGAGGATCTGTCCAAGTTGGCAGTTACTTCTACACAGACATTGCTG ACCTCATGGATGCTGGGAACAAAGCCAGGACAGTGGCAGCTACAAACATGAATGAAAC AAGTAGCCGTTCCACGCTGTGTTTACGATTGTTTTCACCCAGAAGAAACACGATAAT GAGACCAACCTTCCACTGAGAAGGTAGTCAGTAAATCAGCTTGGTGGATCTAGCAG GAAGTGAACGAGCTGATTCAACTGGTGCCAAAGGGAAGTGAATGAAGGAAAGGACAAA TATTAATAAGTCTCTTACAACCTTGGGCAAAGTCATTTACGCTTGGCCGAGGTGAGT AAAAAGAAGAAGAAACAGATTTTATTCCCTACAGGGATTCTGTACTTACTTGGCTCC TTCGAGAAAATTTAGGTGGCAATTCTCGGACTGCAATGGTTGCTGCTGAGCCCCGC GGATATCAACTACGATGAGACTTTGAGCACTCTGAGGTACGAGATCGTGCAAAACAA ATTAATGCAATGCTGTTATCAATGAGGACCCCAATGCCAACTGGTTCTGTAATTAA AGGAGGAGGTGACACGGCTGAAGGACCTTCTCGTGTCTAGGGCCTGGGAGATATTAT TGATGTTGATCCATTGATCGATGATTACTCTGGAAGTGGAAGCAAACGAAAGATTTT CAGAACAATAAGCATAGATACTTGTAGCCTCTGAGAATCAACGCCCTGGCCATTTTT CCACAGCATCCATGGGGTCCCTCACTTTCATCCCCATCTTCTGCTCACTCAGTAGTCA GGTGGGCTTGACGCTCTGTGACCAGTATTCAAGAGAGGATCATGTCTACACCTGGAGGA GAGGAAGCTATTGAACGTTTAAAGGAATCAGAGAAGATCATGTGAGTTGAATGAAA CTTGGGAAGAGAAGCTTCGTAAAACAGAGGCCATCAGAATGGAGAGGGAGGCTTTGTT GGCTGAGATGGGAGTTGCCATTGGGAAGATGGAGGAACCTAGGGGTTTCTCACCT AAAAGACCCACATCTTGTTAACCTCAATGAAGACCCACTAATGTCTGAGTGCCTAC TTTATTACATCAAGATGGAATTACAAGGTTGGCCAAGCAGATGCTGAGCGGGGCCA GGACATAGTGCTGAGCGGGGCTCACATTAAGAAGAGCATGTATCTTCCGGAGTGAG AGAAGCAACAGCGGGGAAGTTATCGTGACCTTAGAGCCCTGTGAGCGCTCAGAAACCT ACGTAAATGGCAAGAGGGTGTCCAGCCTGTTAGCTGCGCTCAGGTAACCGTATCAT CATGGGTAAAACCATGTTTTCCGCTTTAACCACCCGGAACAAGCAGAGCTGAGCGA GAGAAGACTCCTTCTGCTGAGACCCCTCTGAGCCTGTGGACTGGACATTTGCCCAGA GGGAGCTTCTGGAAAAACAAGGAATTGATATGAAACAAGAGATGGAGAAAAGGCTACA GGAAATGGAGATCTTATACAAAAAGGAGAAGGAAGAAGCAGATCTTCTTTGGAGCAG CAGAGACTGGACTATGAGAGTAAATTGCAGGCCTTGCAGAAGCAGGTTGAAACCCGAT CTCTGGCTGCAGAAACAACGAAAGAGGAGGAAGAAGAGGAAGAAGTTCTTGGACACA GCATGAATTTGAGTTGGCCCAATGGGCCTTCCGGAAATGGAAGTCTCATCAGTTTACT TCATTACGGGACTTACTCTGGGGCAATGCCGTGTACCTAAAGGAGGCCAATGCCATCA GTGTGGAACGAAAAAGAAGGTACAGTTTCAAGTTTGTCTGCTGACTGACACAGTGA CTCCCCTTTGCCCTCTGAATTACTTCCCACTGAGATGGAAAAAATCATGAGGACAGG CCTTTCCCTCGCACAGTGGTAGCAGTAGAAGTCCAGGATTGAAAGATGGAGCAACAC ACTATTGGTCTTTGGAGAACTCAAGCAGAGGCTGGATTGATGCGAGAGATGTATGA TAGGGCAGGGGAGATGGCCTCCAGTGCCCAAGACGAAAGCGAAACCACTGTGACTGGC AGCGATCCCTTCTATGATCGGTTCCACTGGTTCAAACCTGTGGGGAGCTCCCCATTT TCCACGGCTGTGTGAACGAGCGCCTTGCCGACCGCACACCTCCCCCACTTTTCCAC GGCCGATTCCGACATCACTGAGCTGGCTGACGAGCAGCAAGATGAGATGGAGGATTTT GATGATGAGGCATTCGTGGATGACGCCGGCTCTGACGAGGGACGGAGGAGGATCAG ATCTCTCAGTGACGGGCATGACCCGTTTACGACCGATCCCCTTGGTTTCAATTTAGT GGAAGGGCATTGTTTACCTGAGCAATCTGCTGTATCCCGTGCCCTGATCCACAGG		

	<p>GTGGCCATCGTCAGTGAGAAAGGTGAAGTGCAGGGGATTTCTGCGTGTGGCTGTACAGG  CCATCGCAGATGAAGAAGCTCCTGATTATGGCTCTGGAATTCGACAGTCAGGAACAGC  TAAATATCTTTTGATAATGAATACTTTAATCAGAGTGACTTTTCGTCTGTTGCAATG  ACTCGTTCTGGTCTGTCCTTGGAGGAGTTGAGGATTGTGGAAGGACAGGGTCAGAGTT  CTGAGGTCATCACTCCTCCAGAAGAAATCAGTCGAATTAAGTACTGTTAGATTGAA  GTCAAGCACTTTGCTGGATGGTAAGATGGTAATGGAAGGGTTTTCTGAAGAGATTGGC  AACCACCTGAAACTGGGCAGTGCCTTCACTTCCGAGTAACAGTGTTGCAGGCCAGTG  GAATCCTCCCAGAGTATGCAGATATCTTCTGTCAGTTCAGCTTTTGTATCGCCATGA  TGAAGCATTCTCCACGGAGCCCCCTCAAAAACAATGGCAGAGGAAGTCCCTGGCCTTT  TATCATGTGCAGAATATTGCAGTGGAGATCACTGAATCATTGTGGATTACATCAAAA  CCAAGCCTATTGTATTGTAAGTCTTTGGGCATTATCAGCAGCACCCACTTCATCTGCA  AGGACAGGAGCTTAACAGTCCGCCTCAGCCGTGCCGCGATTCTTCCCTCCACCCATG  CCACTGTCCAAGCCAGTTCCAGCCACCAAGTTAAACACGATGAGCAAAACCAGCCTTG  GCCAGAGCATGAGCAAGTATGACCTCCTGGTTTGGTTTGGATCAGTGAAGTGGAGCC  TACAGGAGAGTATATCCAGCTGTGGTTGACCACACAGCAGGCTTGCCTTGCCAGGGG  ACATTTTGTCTTCATCAGGGCATCCAGCGAAGGATCACAGTGACCATTTATCCATGAGA  AGGGGAGCGAGCTCCATTGGAAGATGTTCTGTAACGTGGTGGTAGGTGTCGTATTTCG  GAATAAGCCTGAGGTGGATGAAGCTGCAGTTGATGCCATCCTCTCCCTAAATATTATT  TCTGCCAAGTACCTGAAGTCTTCCCAACAATCTAGCAGGACCTTCTACCGCTTTGAGG  CTGTGTGGGATAGCTCTCTGCATAACTCCCTTCTTCTGAACCGAGTGACACCCTATGG  AGAAAAGATCTACATGACCTTGTGGCTTACCTAGAGCTGGATCATTGCATCCAGCCG  GCTGTATCACCAAGGATGTGTGCATGGTCTTCTACTCCCGAGATGCCAAGATCTCAC  CACCACGCTCTCTGCGTAGCCTCTTTGGCAGCGGCTACTCAAAGTACCAGATTCGAA  TCGAGTCACTGGCATTTACGAACTCAGCTTATGCAAAATGTCAGACACAGGTAGTCCA  GGTAAGATGCAGAGAAGGAGAAGAAAAATCTTAGATACGTGAGTGGCATATGTGCGGG  GAGAAGAGAAGCTTAGCAGGCTGGCGGCCCGTGGAGACAGCCTCATCCTTGAGCACCA  GTGGGAGCTGGAGAAGCTGGAAAAAACCAGCCACTTTTGTGCTGCTGCGTGAGAGACTT  GGTGACAGCATCCCCAAATCCCTGAGCGACTCGTTATCCCCCAGCCTCAGCAGTGGGA  CCCTCAGCACCTCCACCAGTATCTCTCTCAGATCTCAACCACTACCTTTGAAAGCGC  CATCACACCTAGCGAGAGCAGTGGCTATGATTGAGGAGACATCGAAGCCTGGTGGAC  CGAGAGAAAGAGCTGGCTACCAAGTGCCTGCAACTTCTCACCACACTTTCAACAGAG  AATTACGCCAGGTGCACGGCAGCGTCAGTGACTGTAAGGTGAGCGATATCTCTCCAAT  TGGACGGGATCCCTCTGAGTCCAGTTTCAGCAGTGCCACCCTCACTCCCTCCTCCACC  TGTCCCTCTCTGGTAGACTCTAGGAGCAACTCTCTGGATCAGAAGACCCAGAGCCA  ATTCGCCGGCCTCTAGTCCCTGCCAGAAATTTGAACAGTTTCAGATTGTCCCAGCTGT  GGAAACACCATAATTTGGCCCCGAGCAGGAAAAAACGAATTTCTCAATCTTGTTCAGAT  ATTGAAGAAATTAGATCAGTGGTCTCTAAGAAAGGATACCTTCATTCAAGGAGCCTC  TTTACAGTAACGGGCTAAACATTTTGTGTGTCGTCGCTCGGCTTATGTCTTCACTA  TAACAGTGACAAAGACCTGTGGAGCGTGAATCATTAACCTGTCCACAGCACAGGTG  GAGTACAGTGAGGACCAGCAGGCCATGGTGAAGACACCAAAACACCTTTGCTGTCTGCA  CAAAGCACCGTGGGGTCTTTTGCAGGCCCTCAATGACAAAGACATGAACGACTGGTT  GTATGCCTTCAACCCACTTCTAGCTGGCACAATACGGAGGTCAAAGCTTTCCCGCAGA  TGCCCGAGCCAGTCGAAATACTAAGTGACTCTGCCGAGTGCCCTCACTCGCCTTCGAG  AGATAAAG</p>
	<div>ORF Start: ATG at 1</div> <div>ORF Stop: TAA at 5416</div>
	<div>SEQ ID NO: 316</div> <div>1805 aa</div> <div>MW at 203184.5kD</div>
NOV31a, CG155962-01 Protein Sequence	<p>MSGASVKVAVRVPFNSRETSKESKCI IQMQGNSTSI INPKNPKEAPKSFSPDYSYWS  HTSPEDPCFASQNRVYNDIGKEMLLHAFEGYNVCI FAYGQTGAGKSYTMMGQESQA  GII PQLCEELFEKINDNCNEEMSYSEVSYMEIYCERVRDLLNPKNGNLRVREHPLL  GPYVEDLSKLAVTSYTDIADLMDAGNKARTVAATNMNETSSRSNAVFTIVFTQKKHDN  ETNLSTEKVVSKI SLVDLAGSERADSTGAKGTRLKEGANINKSLTTLGKVISALAEVS  KKKKKTDFI PYRDSVLTWLLRENLGNSRTAMVAALSPADINDETSLTLRYADRAKQ  IKCNAVINEDPNAKLVLRELKEEVTRLKDLLRAQGLGDI IDVDPLIDDYSGSGSKLKDF  QNNKHRYLLASENQRPGHFSTASMGSLTSSPSSCSLSSQVGLTSVTSIQERIMSTPGG  EEAIERLKESKII AELNETWEEKLRKTEAIRMEREALLAEMGVAIREDGGTLGVFSP  KKTPHLVNLNEDPLMSECLLYYIKDGI TRVGQADAERRQDIVLSGAHIKEEHCFRSE  RSNSGEVIVTLEPCERSETYVNGKRVSPVQLRSGNRI IMGKNHVFRFNHPEQARAER  EKTPSAETPSEPVDTWTFQRELLEKQIDMKQEMEKRLQEMEILYKKEKEEADLLLEQ  QLDYESKLQALQKQVETRSLAAETTEEEEEEEVPWTQHEFELAQWAFRWKSHQFT</p>

SLRDLLWGNAYVLKEANAISVELKKVQFQFVLLTDTLYSPLPPELLPTEMEKTHEDR PFPRTVVAVEVQDLKNGATHYWSLEKLRDLMLREMYDRAGEMASSAQDESETTVTG SDPFYDRFHWFKLVGSSPIFHGCVNERLADRTSPSTFSTADSDITELADEQQDEMEDF DDEAFVDDAGSDAGTEEGSDLFSDGHDPFYDRSPWFILVGRAFVYLSNLLYPVPLIHR VAIVSEKGEVRGFLRVAVQAIADDEEAPDYGSGIRQSGTAKISFDNEYFNQSDFSSVAM TRSGLSLEELRIVEGQGSSEVITPPEEISRINDLLDLKSSSTLLDGKVMMEGFSEEIG NHLKLGSFTFRVTVLQASGILPEYADIFCQFSFLHRHDEAFSTEPLKNNGRGSPLAF YHVQNIABITESFVDYIKTKPIVFEVFGHYQQHPLHLQGQELNSPPQPCRRFFPPPM PLSKPVPATKLNMTSKTSLGQSMKYDLLVWFIEISELEPTGEYIPAVVDHTAGLPCQG TFLHQQGIQRRITVTIIEKGSSELHWKDVRELTVGGRIIRNKPEVDEAAVDAILSLNII SAKYLKSSHNSRTPYRFEAVWDSLSLHNSLLNLRVTPYGEKIYMTLSAYLELDHICIQP AVITKDVCMVFYSRDAKISPPRSLRSLFGSGYSKSPDSNRVTGIYELSLCKMSDTGSP GKMQRRRRKILDTSVAYVRGEENLAGWRPRGDSLILEHQWELEKLEKTRHFLRLRERL GDSIPKSLSDSLSPSLSSGTLSTSTSISSQISTTTFESAITPSESSGYDSGDIESLVD REKELATKCLQLLTHTFNREFSQVHGSVSDCKVSDISPIGRDPSESSFSSATLTPSST CPSLVDSRSNSLDQKTPANSRASSPCPEFEQFQIVPAVETPYLARAGKNEFLNLVPD IEEIRSVVSKKGYLHFKEPLYSNWAKHFVVVRFPYVFIYNSDKDPVERGIINLSTAQV EYSEDQQAMVKTPTNTFAVCTKHRGVLLQALNDKMDNDWLYAFNPLLAGTIRRSKLSRR CPSQSKY.
--

Further analysis of the NOV31a protein yielded the following properties shown in Table 31B.

Table 31B. Protein Sequence Properties NOV31a	
PSort analysis:	0.5985 probability located in mitochondrial matrix space; 0.4900 probability located in nucleus; 0.3052 probability located in mitochondrial inner membrane; 0.3052 probability located in mitochondrial intermembrane space
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV31a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 31C.

5

Table 31C. Geneseq Results for NOV31a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV31a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAB36227	Human kinesin-like protein HKLP SEQ ID NO: 4 - Homo sapiens, 1816 aa. [WO200063375-A1, 26-OCT-2000]	1..1805 1..1816	1797/1821 (98%) 1800/1821 (98%)	0.0
ABB07867	Human kinesin-associated protein having motor domain - Homo sapiens, 1823 aa. [WO200226965-A1, 04-APR-2002]	1..1804 1..1816	1785/1820 (98%) 1790/1820 (98%)	0.0
ABB07866	Human kinesin-associated protein	430..1805	1370/1385 (98%)	0.0

	lacking motor domain - Homo sapiens, 1381 aa. [WO200226965-A1, 04-APR-2002]	1..1381	1372/1385 (98%)	
AAU28137	Novel human secretory protein, Seq ID No 306 - Homo sapiens, 1381 aa. [WO200166689-A2, 13-SEP-2001]	430..1805 1..1381	1370/1385 (98%) 1372/1385 (98%)	0.0
AAU28325	Novel human secretory protein, Seq ID No 682 - Homo sapiens, 1374 aa. [WO200166689-A2, 13-SEP-2001]	439..1805 3..1374	1355/1376 (98%) 1360/1376 (98%)	0.0

In a BLAST search of public sequence databases, the NOV31a protein was found to have homology to the proteins shown in the BLASTP data in Table 31D.

Table 31D. Public BLASTP Results for NOV31a				
Protein Accession Number	Protein/Organism/Length	NOV31a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
O60333	Kinesin-like protein KIF1B (Klp) - Homo sapiens (Human), 1816 aa.	1..1805 1..1816	1783/1821 (97%) 1791/1821 (97%)	0.0
Q60575	Kinesin-like protein KIF1B - Mus musculus (Mouse), 1816 aa.	1..1805 1..1816	1745/1821 (95%) 1783/1821 (97%)	0.0
Q8R524	Kinesin-family protein 1Bp204 - Rattus norvegicus (Rat), 1816 aa.	1..1805 1..1816	1741/1821 (95%) 1779/1821 (97%)	0.0
Q96Q94	Kinesin-related protein - Homo sapiens (Human), 1388 aa.	430..1804 1..1381	1359/1384 (98%) 1363/1384 (98%)	0.0
O88658	Kinesin-like protein KIF1B - Rattus norvegicus (Rat), 689 aa (fragment).	1..700 1..689	657/704 (93%) 668/704 (94%)	0.0

PFam analysis predicts that the NOV31a protein contains the domains shown in the Table 31E.

Table 31E. Domain Analysis of NOV31a			
Pfam Domain	NOV31a Match Region	Identities/ Similarities for the Matched Region	Expect Value

kinesin	11..378	183/418 (44%) 323/418 (77%)	6.7e-188
FHA	550..621	22/85 (26%) 55/85 (65%)	1.6e-14
PH	1690..1787	28/98 (29%) 78/98 (80%)	4.6e-18

**Example 32.**

The NOV32 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 32A.

Table 32A. NOV32 Sequence Analysis			
	SEQ ID NO: 317	3120 bp	
NOV32a, CG157477-01 DNA Sequence	GGAGGCCCGAGCGCGCCACCTGAGCCCCGCGCTGGCGCCATGGCGGAGCAGGAGA GCCTGGAATTCGGCAAGGCAGACTTCGTGCTGATGGACACCGTCTCCATGCCCCGAGTT CATGGCCAACCTCAGGCTCAGATTTGAAAAAGGGCGCATCTATACGTTTCATTGGAGAA GTCGTCGTTTCTGTGAACCCCTTACAAGTTGTTGAACATCTATGGAAGAGACACAATTG AGCAGTATAAAGGCCGTGAGCTGTATGAGAGACCCCTCACCTTTTTGCTATTGCGGA TGCTGCTTACAAGGCTATGAAGAGGCGATCAAAAGACACTTGATTGTGATATCAGGG GAAAGTGGAGCTGGTAAAAACGGAAGCCAGTAAGTACATTATGCAGTATATTGCGGCCA TCACCAACCCAGTCAGAGAGCAGAGGTTGAAAGAGTGAAGAATATGTTGCTTAAGTC CAACTGTGTTTTGGAAGCTTTTGGAAATGCCAAACCAACCGTAATGACAACTCAAGC AGGTTTGGAAAATACATGGATATCAACTTTGACTTCAAGGGTGACCCATTGTTGGTGGGC ATATCAATAACTACTTACTAGAAAAGTCTCGAGTGATTGTGCAACAGCCAGGAGAAAG AAGCTTTCATTCTTTCTATCAGCTACTCCAAGGAGGTTTCAAGACAAATGCTACGCTCT CTACATCTCCAGAAATCCCTTTTCACTCCTACAATATATTCATGTGGGAGCTCAATTAA AGTCTTCTATCAATGATGCTGCCGAATTCAGAGTTGTTGCTGATGCCATGAAAGTCAT TGGCTTCAAACCTGAGGAGATCCAAACAGTGTATAAGATTTTGGCTGCTATTCTGCAC TTGGGAAATTTAAATTTGTAGTAGATGGTGACACGCCTCTATTGAGAATGGCAAAG TAGTATCTATCATAGCAGAATTGCTCTCTACTAAGACAGATATGTTTGAGAAAGCCCT TCTTTACCGGACTGTGGCCACAGGCCGTGACATCATTGACAAGCAGCACACAGAACAA GAGGCCAGCTACGGCAGAGACGCCTTTGCCAAGGCAATATATGAGCGCCTTTTTTGT GGATCGTTACTCGCATCAATGATATTATTGAGGTCAAGAACTATGACACCACAATCCA TGGGAAGAACACTGTTATTGGTGTCTTGGATATCTATGGCTTTGAAATCTTTGACAAC AACAGTTTGAACAATTCTGTATCAATTACTGCAATGAGAACTGCAGCAGCTATTTA TTCAGCTGGTTCTGAAGCAAGAACAAGAGGAATACCAGCGGGAAGGGATCCCCCTGGAA ACATATTGACTACTTCAACAATCAGATCATTGTTGACCTCGTGGAGCAACAGCACAAA GGGATCATTGCAATCCTTGATGATGCTTGATGAATGTGCGCAAAGTCACCGATGAAA TGTTTCTTGAAGCACTTAACAGTAAATTGGGCAAAACGCCCATTTTCCAGCCGAAA GCTCTGTGCCTCAGACAAAATTCTGGAGTTTGATCGAGATTTTCAATTCGACATTAT GCAGGCGATGTAGTCTATTCTGTCTATTGGTTTATTGACAAAAATAAGATACTTTAT TTCAAGATTTCAAGCGCCTTATGTATAACAGTTCAAATCCTGTGCTCAAGAATATGTG GCCTGAAGGCAAACTGAGCATTACAGAGGTGACCAAGCGACCTCTGACTGCTGCTACC TTGTTTAAGAATTCTATGATTGCTCTAGTAGACAACCTTGATCAAAAGGAACCATATT ACGTTTCGTGTCATCAAAACCAATGACAAGAAATCTCCACAGATATTTGATGATGAACG CTGCCGGCACCAAGTAGAATATCTTGGACTACTGGAATGTGAGAGTGCCTCGGGCA GGATTTGCCTTCCGCCAGACATACGAGAAGTTTCTTCAAGGTATAAGATGATCTCTG AATTCACCTGGCCCAACCATGACCTTCTTCAAGACAAAGAGGCTGTCAAGAACTAAT TGAACGGTGTGGTTTTTCAAGATGATGTAGCTTATGGGAAGACCAAAATTTTCATTGCA ACACCCCAACATGTTTACCTTGAAGAACTCCGTGCCCGAGATGCTCATAAGGATTG TCCTCTTTCTACAAAAGGTGTGGCGGGGCACCCTGGCCCGCATGCGGTACAAAAGAAC CAAGGCAGCTCTGACAATAATCAGGTACTACCGGCGCTACAAAGTGAAGTGCATCATC CACGAGGTGGCCAGACGCTTCCATGGCGTCAAGACCATGCGGAGACTACGGGAAGCACG TGAAGTGGCCAAGCCCTCTAAAGTTCTTCGCCGTTTTGAGGAGGCCCTGCAGACGAT		

	TTTCAATAGATGGAGAGCATCCCAGCTCATCAAGAGCATTCCGGCCTCAGACCTGCC CAGGTCAGGGCAAAGGTTGCAGCCGTGGAAATGTTGAAGGGTCAAAGGGCTGACCTCG GGCTCCAGAGGGCCTGGGAGGGCAACTATCTTGCTTCAAAGCCAGATACACCTCAGAC CTCAGGCACTTTTGTCCCTGTTGCTAATGAATTGAAACGGAAGGACAAATACATGAAT GTCCTCTTTTCTGTACGCTCCGTAAGGTAAATCGATTTAGTAAGGTGGAAGACAGAG CAATTTTTGTCACTGACCGTCACCTGTATAAAATGGATCCCACTAAACAGTACAAGGT GATGAAGACTATCCCTCTATACAATTTGACTGGTCTGAGTGTCTCCAATGGAAAGGAC CAACTTGTAGTGTTCATACGAAAGACAACAAAGACCTCATTGTCTGCCTCTTCAGCA AACAGCCAACCCATGAGAGTCGAATTGGAGAACTTGTGGAGTGTGGTGAATCATT CAAGAGTGAGAAGCGCCACCTTCAAGTGAACGTACCAACCCAGTACAGTGCAGCCTG CACGGGAAGAAGTGACCGTCTCCGTGGAGACGCGGCTCAACCAGCCCCAGCCCGACT TCACCAAGAATCGCTCGGGCTTCATCTCAGCGTGCCCGGGAAGTACGCCCGCGGGA GGCCTGGCCCGGAGCCCGGCCACACTCCGAGTCTGGGTCCAGTC		
	ORF Start: ATG at 43		ORF Stop: TGA at 3061
	SEQ ID NO: 318	1006 aa	MW. at 116201.0kD
NOV32a, CG157477-01 Protein Sequence	MAEQESLEFGKADFVIMDTVSMPEFMANLRLRFEKGRIYTFIGEVVSVNPYKLLNIY GRDTIEQYKGRELYERPPHLFAIADAAYKAMKRRSKDTCIVISGESGAGKTEASKYIM QYIAAITNPSQRAEVERVKMMLLSNCVLEAFGNAKTNRNDNSSRFGKYMDINFDFKG DPIGGHINNYLLEKSRVIVQPGERSFHSFYQLLQGGSEQMLRSLHLQKSLSSYNIH VGAQLKSSINDAAEFRVADAMKVIQFKPEEIQTVYKILAAIHLHLGNLKFVVDGDTPL IENGKVVSI IAE LLSTK TDMVEKALLYRTVATGRDIIDKQHTEQEASYGRDAFAKAIY ERLFCWIVTRINDIIEVKNYDTTIHGKNTVIGVLDIYGFEI FDNNSFEQFCINYCNEK LQQLFIQLVLKQEQEYQREGIPWKHIDYFNNQIIVDLVEQKHGIIAILDDACMNVG KVTDEMFLALNSKLKHAHFSSRKL CASDKILEFDRDFRIRHYAGDVVSVIGFIDK NKDTLFDQDFKRLMYNSSNPVLKNMWPEGKLSITEVTKRPLTAATLFKNSMIALVDNLA SKEPYVRCIKPNDKSPQIFDDERCRHQVEYLGLLENVRVRAGFAFRQTYEKLHR YKMISEFTWPNHDLPSDKEAVKKLIERCGFQDDVAYGKTKI FIRTPRTLFTLEELRAQ MLIRIVLFLQKVWRGTLARMRYKRTKAALTIIRYRRYKVKSYTHEVARRFHGVKTMR DYGKHKVWPSPPKVLRREFEALQTFNRWRASQLIKSIPASDLPQVRAKVA AVEMLKG QRADLGLQRAWEGNYLASKPDTPTSGTFVPVANELKRKDKYMNVLFSCHVRKVNRF KVEDRAIFVTDRLHYKMDPTKQYKVMKTIPLYNLTGLSVSNGKDQLVVFHTKDNKDLI VCLFSKQPTHE SRIGELVGVLVNHFHFKSEKRHLQVNVNTPVQCSLHGKCTVSVETRLN QPQPDFTKNRSGFILSVPCN		

Further analysis of the NOV32a protein yielded the following properties shown in Table 32B.

Table 32B. Protein Sequence Properties NOV32a	
PSort analysis:	0.7600 probability located in nucleus; 0.3760 probability located in microbody (peroxisome); 0.1000 probability located in mitochondrial matrix space; 0.1000 probability located in lysosome (lumen)
SignalP analysis:	No Known Signal Sequence Predicted

- 5 A search of the NOV32a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 32C.

Table 32C. Geneseq Results for NOV32a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV32a Residues/	Identities/ Similarities for	Expect Value

		<b>Match Residues</b>	<b>the Matched Region</b>	
AAM80123	Human protein SEQ ID NO 3769 - Homo sapiens, 764 aa. [WO200157190-A2, 09-AUG-2001]	243..1006 1..764	764/764 (100%) 764/764 (100%)	0.0
AAM79139	Human protein SEQ ID NO 1801 - Homo sapiens, 753 aa. [WO200157190-A2, 09-AUG-2001]	254..1006 1..753	752/753 (99%) 752/753 (99%)	0.0
ABG16605	Novel human diagnostic protein #16596 - Homo sapiens, 674 aa. [WO200175067-A2, 11-OCT-2001]	333..1006 1..674	670/674 (99%) 671/674 (99%)	0.0
AAU23125	Novel human enzyme polypeptide #211 - Homo sapiens, 1026 aa. [WO200155301-A2, 02-AUG-2001]	1..1004 9..1024	611/1016 (60%) 784/1016 (77%)	0.0
AAU23128	Novel human enzyme polypeptide #214 - Homo sapiens, 909 aa. [WO200155301-A2, 02-AUG-2001]	1..841 9..861	532/853 (62%) 676/853 (78%)	0.0

In a BLAST search of public sequence databases, the NOV32a protein was found to have homology to the proteins shown in the BLASTP data in Table 32D.

<b>Table 32D. Public BLASTP Results for NOV32a</b>				
<b>Protein Accession Number</b>	<b>Protein/Organism/Length</b>	<b>NOV32a Residues/ Match Residues</b>	<b>Identities/ Similarities for the Matched Portion</b>	<b>Expect Value</b>
Q63357	Myosin I - Rattus norvegicus (Rat), 1006 aa.	1..1006 1..1006	985/1006 (97%) 998/1006 (98%)	0.0
A53933	myosin I myr 4 - rat, 1006 aa.	1..1006 1..1006	983/1006 (97%) 996/1006 (98%)	0.0
O94832	KIAA0727 protein - Homo sapiens (Human), 674 aa (fragment).	333..1006 1..674	674/674 (100%) 674/674 (100%)	0.0
Q23978	Myosin IA (MIA) (Brush border myosin IA) (BBMIA) - Drosophila melanogaster (Fruit fly), 1011 aa.	8..1004 6..1006	542/1004 (53%) 706/1004 (69%)	0.0
S45573	myosin IA - fruit fly (Drosophila melanogaster), 1011 aa.	8..1004 6..1006	541/1004 (53%) 704/1004 (69%)	0.0

PFam analysis predicts that the NOV32a protein contains the domains shown in the Table 32E.

Table 32E. Domain Analysis of NOV32a			
Pfam Domain	NOV32a Match Region	Identities/ Similarities for the Matched Region	Expect Value
myosin_head	13..682	314/743 (42%) 544/743 (73%)	0
IQ	699..719	10/21 (48%) 16/21 (76%)	0.0053

### Example 33.

5 The NOV33 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 33A.

Table 33A. NOV33 Sequence Analysis			
	SEQ ID NO: 319	3921 bp	
NOV33a, CG157486-01 DNA Sequence	CAGAAGTTGCGCGCAGGCCGCGGGCGGGAGCGGACACCGAGGCCGCGTGCAGGCCGT GCGGGTGTGCGGGAGCCGGGCTCGGGGGGATCGGACCGAGAGCGAGAAGCGCGGCATG GAGCTCCAGGCAGCCCGCGCTGCTTCGCCCTGCTGTGGGGCTGTGCGCTGGCCGCGG CCGCGGCGGCGCAGGGCAAGGAAGTGGTACTGCTGGACTTTGCTGCAGCTGGAGGGGA GCTCGGCTGGCTCACACACCCGTATGGCAAAGGGTGGGACCTGATGCAGAACATCATG AATGACATGCCGATCTACATGTACTCCGTGTGCAACGTGATGTCTGGCGACCAGGACA ACTGGCTCCGCACCAACTGGGTGTACCGAGGAGAGGGCTGAGCGTATCTTCATTGAGCT CAAGTTTACTGTACGTGACTGCAACAGCTTCCCTGGTGGCGCCAGCTCCTGCAAGGAG ACTTTCAACCTCTACTATGCCGAGTCGGACCTGGACTACGGCACCAACTTCCAGAAGC GCCTGTTACCAAGATTGACACCATTTGCGCCCGATGAGATCACCGTCAGCAGCGACTT CGAGGCACGCCACGTGAAGCTGAACGTGGAGGAGCGCTCCGTGGGGCCGCTCACCCGC AAAGGCTTCTACCTGGCCTTCCAGGATATCGGTGCCCTGTGTGGCGCTGCTCTCCGTCC GTGTCTACTACAAGAAGTGCCCCGAGCTGCTGCAGGGCCTGGCCCACTTCCCTGAGAC CATCGCCGGCTCTGATGCACCTTCCCTGGCCACTGTGGCCGCGACCTGTGTGGACCAT GCCGTGGTGCCACCGGGGGGTGAAGAGCCCCGTATGCACTGTGCAGTGGATGGCGAGT GGCTGGTGCCATTGGGCAGTGCCTGTGCCAGGCAGGCTACGAGAAGGTGGAGGATGC CTGCCAGGCCTGCTCGCTGGATTTTTTAAGTTTGAGGCATCTGAGAGCCCCGCTTG GAGTGCCCTGAGCACACGCTGCCATCCCCTGAGGGTGCCACCTCCTGCGAGTGTGAGG AAGGCTTCTTCCGGGCACCTCAGGACCCAGCGTCGATGCCTTGACACGACCCCCCTC CGCCCCACACTACCTCACAGCCGTGGGCATGGGTGCCAAGGTGGAGCTGCGCTGGACG CCCCCCAGGACAGCGGGGGCCGCGAGGACATTGTCTACAGCGTCACCTGCGAACAGT GCTGGCCCCAGTCTGGGGAATGCGGGCCGTGTGAGGCCAGTGTGCGCTACTCGGAGCC TCCTCACGGACTGACCCGACCAAGTGTGACAGTGAGCGACCTGGAGCCCCACATGAAC TACACCTTACCGTGGAGGCCCGCAATGGCGTCTCAGGCCTGGTAACAGCCGACGT TCCGTACTGCCAGTGTGAGCATCAACCAGACAGAGCCCCCAAGGTGAGGCTGGAGGG CCGCAGCACCACTCGCTTAGCGTCTCCTGGAGCATCCCCCGCCGAGCAGAGCCGA GTGTGGAAGTACGAGGTCACTTACCGCAAGAAGGGAGACTCCAACAGCTACAATGTGC GCCGACCCAGGGTTTCTCCGTGACCCTGGACGACCTGGCCCCAGACACCACCTACCT GGTCCAGGTGCAGGCACTGACGAGGAGGGCCAGGGGGCCGGCAGCAAGGTGCACGAA TTCCAGACGCTGTCCCGGAGGGATCTGGCAACTTGGCGGTGATTGGCGGCGTGGCTG TCGGTGTGCTCCTGCTTCTGGTGTGCGAGGAGTTGGCTTCTTTATCCACCGCAGGAG GAAGAACCAGCGTGCCCGCCAGTCCCCGGAGGACGTTTACTTCTCAAGTCAGAACAA CTGAAGCCCTGAAGACATACGTGGACCCCCACACATATGAGGACCCCAACCAGGCTG TGTTGAAGTTCACCTACCGAGATCCATCCATCCTGTGTCACTCGCAGAAAGGTGATCGG		

	AGCAGGAGAGTTTGGGGAGGTGTACAAGGGCATGCTGAAGACATCCTCGGGGAAGAAG GAGGTGCCGGTGGCCATCAAGACGCTGAAAGCCGGCTACACAGAGAAGCAGCGAGTGG ACTTCCTCGGCGAGGCCGGCATCATGGGCCAGTTCAGCCACCACAACATCATCCGCCCT AGAGGGCGTCATCTCCAAATACAAGCCCATGATGATCATCACTGAGTACATGGAGAAT GGGGCCCTGGACAAGTTCCTTCGGGAGAAGGATGGCGAGTTTCAGCGTGTGCGAGTGG TGGGCATGCTGCGGGGCATCGCAGCTGGCATGAAGTACCTGGCCAACATGAACATATGT GCACCGTGACCTGGCTGCCCGCAACATCCTCGTCAACAGCAACCTGGTCTGCAAGGTG TCTGACTTTGGCCTGTCCCGCGTGTGGAGGACGACCCCGAGGCCACCTACACCACCA GTGGCGGCAAGATCCCCATCCGCTGGACCGCCCCGGAGGCCATTTCTACCGGAAGTT CACCTCTGCCAGCGACGTGTGGAGCTTTGGCATTGTATGTGGGAGGTGATGACCTAT GGCGAGCGGCCCTACTGGGAGTTGTCCAACCACGAGGTGATGAAAGCCATCAATGATG GCTTCCGGCTCCCCACACCCATGGACTGCCCCCTCCGCCATCTACCAGCTCATGATGCA GTGCTGGCAGCAGGAGCGTGGCCGCCGCCCAAGTTCGCTGACATCGTCAGCATCCTG GACAAGCTCATTCGTGCCCTGACTCCCTCAAGACCTGGCTGACTTTGACCCCCGCG TGTCTATCCGGCTCCCCAGCAGAGCGGCTCGGAGGGGTGCCCTTCCGCACGGTGTG CGAGTGGCTGGAGTCCATCAAGATGCAGCAGTATACGGAGCACTTCATGGCGGCCGGC TACACTGCCATCGAGAAGGTGGTGCAGATGACCAACGACGACATCAAGAGGATTGGGG TGCGGCTGCCCGGCCACCAGAAGCGCATCGCCTACAGCCTGTGGGACTCAAGGACCA GGTGAACACTGTGGGGATCCCCATCTGAGCCTCGACAGGGCCTGGAGCCCCATCGGCC AAGAATACTTGAAGAAACAGAGTGGCCTCCCTGCTGTGCCATGCTGGGCCACTGGGGA CTTTATTTATTTCTAGTTCTTTCTCCCCCTGCAACTTCCGCTGAGGGGTCTCGGATG ACACCCTGGCCTGAACTGAGGAGATGACAGGGATGCTGGGCTGGGCCCTCTTTCCCT GCGAGACGCACACAGCTGAGCACTTAGCAGGCACCGCCACGTCCAGCATCCCTGGAG CAGGAGCCCCGCCACAGCCTTCGGACAGACATATGGGATATTTCCAAGCCGACCTTCC CTCCGCCCTTCTCCACATGAGGCCATCTCAGGAGATGGAGGGCTTGGCCCCAGCGCCAA GTAAACAGGGTACCTCAAGCCCCATTTCCTCACACTAAGAGGGCAGACTGTGAACCTG ACTGGGTGAGACCCAAAGCGGTCCCTGTCCCTCTAGTGCCCTCTTTAGACCCCTCGGGC CCCATCCTCATCCCTGACTGGCCAAACCTTGCTTTCTGGGCTTTGCAAGATGCTT GGTTGTGTGAGGTTTTTAAATATATATTTTGTACTTTGTGGAGAGAATGTGTGTGTG TGGCAGGGGGCCCCGCCAGGGCTGGGGACAGAGGGTGTCAAACATTCGTGAGCTGGGG ACTCAGGACCGGTGCTGCAGGAGTGTCTGCCATGCCCCAGTCGGCCCCATCTCTC ATCCTTTGGATAAGTTTCTATTCTGTCACTGTAAAGATTGTGTTTGTGGACATT TTTTTCGAATCTTAATTTATTATTTTATATTTATTGTTAGAAAATGACTTATTT CTGCTCTGGAATAAAGTTCAGATGATTCAAACCG
	ORF Start: ATG at 114
	ORF Stop: TGA at 3042
	SEQ ID NO: 320
	976 aa
	MW at 108265.3kD
NOV33a, CG157486-01. Protein Sequence	MELQAARACFALLWGCALAAAAAQKEVVLDDFAAAGGELGWLTHPYKGWDLQMNI MNDMPIYMSVCNVMMSGDQDNWLRNTNWVYRGEAERIFIELKFTVRDCNSFPGGASSCK ETFNLYAESDLGYTNFQKRLFTKIDTIAPDEITVSSDFEARHVKLNVEERSVGPLT RKGFYLAQDIGACVALLSVRVYKCKPELLQGLAHFPETIAGSDAPSLATVAGTCVD HAVVPPGGEEPMMHCAVDGEWLVPQGCLCQAGYEKVEDACQACSPGFFKFEASESPC LECPHTLPSPEGATSCCEEGFFRAPQDPASMPCTRPPSAPHYLTAVGMGAKVELRW TPPQDSGGREDIVSVTCEQCWPESGECGPCEASVRYSEPPHGLTRTSVTVSDLEPHM NYTFTVEARNGVSGLVTSRSFRITASVSINQTEPPKVRLEGRSTTSLSVSWSIPPPQQS RVWKYEVTYRKKGDSNSYNVRRTEGFSVTLDDLPDTTYLVQVQALTQEQGAGSKVH EFQTLSPGSGNLAVIGGVAVGVVLLVLAVGVFFIHRRRKNQRRARQSPEDVYFSKSE QLKPLKTYVDPHTYEDPNQAVLKFTTEIHPSCVTRQKVI GAGEFGEVYKGMKLTSSGK KEVPVAIKTLKAGYTEKQRVDFLGEAGIMGQFHHNIIRLEGVISKYKPMMIITEYME NGALDKFLREKDGFSVLQVLGMLRGIAAGMKYLANMNYVHRLAARNILVNSNLVCK VSDFGLSRVLEDDPEATYTTSGGKIPIRWTAPEAISYRKFTSASDVWSFGIVMWEVMT YGERPYWELSNHEVMKAINDFRLPTPMDCPSAIYQLMMQCWQERARRPKFADIVSI LDKLI RAPDSLKTADFDPRVIRLPSTSGSEGVFPRTVSEWLESIKMQQYTEHFMMAA GYTAIEKVVMNTDDIKRIGVRLPGHQKRIAYSLGLKLDQVNTVGIPI

Further analysis of the NOV33a protein yielded the following properties shown in Table 33B.

<b>Table 33B. Protein Sequence Properties NOV33a</b>	
<b>PSort analysis:</b>	0.4600 probability located in plasma membrane; 0.1000 probability located in endoplasmic reticulum (membrane); 0.1000 probability located in endoplasmic reticulum (lumen); 0.1000 probability located in outside
<b>SignalP analysis:</b>	Cleavage site between residues 24 and 25

A search of the NOV33a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 33C.

<b>Table 33C. Geneseq Results for NOV33a</b>				
<b>Geneseq Identifier</b>	<b>Protein/Organism/Length [Patent #, Date]</b>	<b>NOV33a Residues/ Match Residues</b>	<b>Identities/ Similarities for the Matched Region</b>	<b>Expect Value</b>
AAR85090	EPH-like receptor protein tyrosine kinase HEK7 - Homo sapiens, 991 aa. [WO9528484-A1, 26-OCT-1995]	11..976 14..991	524/984 (53%) 680/984 (68%)	0.0
AAR85092	EPH-like receptor protein tyrosine kinase HEK11 - Homo sapiens, 998 aa. [WO9528484-A1, 26-OCT-1995]	13..969 16..988	504/979 (51%) 659/979 (66%)	0.0
AAW03421	Mouse developmental kinase 1 - Mus sp, 998 aa. [WO9621013-A1, 11-JUL-1996]	9..969 14..988	505/982 (51%) 660/982 (66%)	0.0
AAW83147	Rat receptor tyrosine kinase Ehk-1 - Rattus sp, 1005 aa. [US5843749-A, 01-DEC-1998]	13..940 42..1003	503/969 (51%) 654/969 (66%)	0.0
AAB08665	Amino acid sequence of a human EphA3 HLA class II-binding peptide - Homo sapiens, 983 aa. [WO200050589-A1, 31-AUG-2000]	28..976 29..983	499/964 (51%) 652/964 (66%)	0.0

5. In a BLAST search of public sequence databases, the NOV33a protein was found to have homology to the proteins shown in the BLASTP data in Table 33D.

<b>Table 33D. Public BLASTP Results for NOV33a</b>				
<b>Protein Accession Number</b>	<b>Protein/Organism/Length</b>	<b>NOV33a Residues/ Match</b>	<b>Identities/ Similarities for the Matched</b>	<b>Expect Value</b>

		Residues	Portion	
AAH37166	EphA2 - Homo sapiens (Human), 976 aa.	1..976 1..976	976/976 (100%) 976/976 (100%)	0.0
P29317	Ephrin type-A receptor 2 precursor (EC 2.7.1.112) (Tyrosine-protein kinase receptor ECK) (Epithelial cell kinase) - Homo sapiens (Human), 976 aa.	1..976 1..976	972/976 (99%) 972/976 (99%)	0.0
Q03145	Ephrin type-A receptor 2 precursor (EC 2.7.1.112) (Tyrosine-protein kinase receptor ECK) (Epithelial cell kinase) (MPK-5) (SEK-2) - Mus musculus (Mouse), 977 aa.	1..976 1..977	905/978 (92%) 931/978 (94%)	0.0
I48974	receptor-protein tyrosine kinase - mouse, 975 aa.	1..976 1..975	886/978 (90%) 916/978 (93%)	0.0
Q9PWR5	Eph receptor tyrosine kinase precursor - Xenopus laevis (African clawed frog), 977 aa.	25..976 24..977	690/957 (72%) 798/957 (83%)	0.0

PFam analysis predicts that the NOV33a protein contains the domains shown in the Table 33E.

Table 33E. Domain Analysis of NOV33a			
Pfam Domain	NOV33a Match Region	Identities/ Similarities for the Matched Region	Expect Value
EPH_lbd	28..201	103/178 (58%) 167/178 (94%)	2.4e-126
fn3	329..424	29/98 (30%) 72/98 (73%)	4.1e-12
fn3	436..519	32/87 (37%) 67/87 (77%)	2.3e-20
pkinase	613..868	82/292 (28%) 204/292 (70%)	1.7e-75
SAM	902..966	30/68 (44%) 58/68 (85%)	7.1e-26

#### Example 34.

5 The NOV34 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 34A.

Table 34A. NOV34 Sequence Analysis

	SEQ ID NO: 321	14399 bp
NOV34a, CG157505-01 DNA Sequence	ATGGCGAACGTGCAGGTCGCCGTGCGGGTCCGGCCGCTCAGCAAGAGGGAGACCAAAG AAGGGGGAAGAATTATTGTGGAAGTTGATGGCAAAGTGGCAAAAATCAGGAATTTAA GGTAGACAAATCGACCAGATGGCTTTGGGGACTCCGGGAGAAGGTTATGGCATTGGC TTTGATTACTGCTACTGGTCAGTCAACCCAGAGGATCCCCAGTATGCATCTCAAGATG TGGTATTCCAGGATTTAGGGATGGAAGTACTGTCTGGAGTTGCCAAAGGCTATAACAT ATGCCTTTTGTCTTATGGACAGACAGGCTCTGGGAAGACATATACCATGCTGGGGACC CCAGCCTCTGTTGGGTTGACACCACGGATATGTGAGGGTCTCTTCGTGAGGGAGAAAG ACTGTGCCTCACTGCCTTCCTCCTGTAGGATAAAAAGTAAGTTTTCTAGAAATCTATAA TGAACGGGTGCGGGATCTGTTGAAGCAATCTGGTCAAAAAAAGTCTTATACCTTGCGG GTCAGGGAGCATCCAGAGATGGGGCCCTATGTACAAGGTTTATCTCAACATGTAGTTA CCAATTATAAGCAAGTAATCCAACCTCTGGAGGAGGGAATTGCAACAGGATCACAGC AGCCACCCATGTTTCATGAGGCCAGCAGCAGATCCACGCCATTTTCACGATCCACTAC ACGCAGGCAATCCTGGAGAACAACCTCCCTTCTGAAATGGCTAGCAAGATCAACCTTG TGGACCTAGCAGGCAGCGAAAGAGCAGATCCAGTTACTGTAAGGACCGCATTTGCTGA AGGAGCCAATATCAACAAGTCCCTTGTGACTCTAGGAATTGTCTATCTCCACCTTAGCC CAGAACTCCCAAGTTTTTCAGCAGCTGCCAGAGCCTCAACAGCTCAGTCAGCAATGGTG GTGACAGTGGGATCCTTAGCTCTCCTTCTGGGACCAGCAGTGGAGGGGACCTCCCG AAGGCAGTCTTATATCCCATACCGAGACTCTGTGTTGACCTGGCTGCTGAAGGACAGC CTTGGAGGCAACTCTAAAACCATCATGGTTGCCAGTGTGTCTCCTGCACACACTAGCT ACAGTGAGACCATGAGCACACTGAGATATGCATCCAGTGCCAAAAACATTATCAACAA GCCACGAGTAAATGAGGATGCAAACCTAAAACCTGATTAGAGAACTCAGAGAAGAGATT GAAAGACTGAAAGCCCTGCTGCTGAGCTTTGAACTGAGAAATTCAGTTTCATTGAGTG ATGAAAACCTGAAGGAGCTGGTTCTCCAAAATGAATTGAAGATAGACCAGCTGACTAA AGACTGGACCCAGAAGTGAATGATTGGCAGGCCCTCATGGAGCATTACAGTGTGGAC ATCAACAGGAGGAGGGCTGGGGTGGTCATCGACTCCAGCCTGCACACACTTGATGGCCT TGGAGGATGATGTGCTCAGCACAGGTGTTGTGCTCTATCATCTCAAGGAAGGGACAAC AAAAATAGGAAGGATTGACTCAGACCAGGAACAGGACATTGTCTGCAGGGTCAGTGG ATTGAGAGAGACCACCTGCACTATCACCAGTGCCTGTGGTGTAGTTGTTCTACGACCTG CCCGTGGGGCCCGCTGTACAGTCAATGGCCGGGAGGTCAGTCCCTCCTGCCGTCTGAC TCAAGGAGCTGTATATAACCTTGGGGAGGCCACAGAAGTTCCGATTCAACACCCAGCA GAGGCTGCTGTCTGCGGCAGCGAAGGCAGGTTGGAGAGGCTGCTGCTGGTCTGGCT CGTTGGAGTGGCTGGATTGATGGAGATCTCGCTGCCTCCCGGCTGGGTCTCTCCCC TTTGCTTTTGAAGGAAAGGAGAGCGCTTGAAGAGCAATGTGACGAGGACCATCAGACA CCGAGGGATGGAGAGACATCCACAGGGCCAGATTTCAGCAGCAGCAGAGCTACGTAG AGGATTTGAGGCATCAAACTCCTAGCAGAAGAGATTGAGCTGCCAAGGAACCTGGAATT TGACCAAGCTTGGATTAGCCAGCAGATTAAAGAAAACAGCAGTGTCTGCTCAGAGAA GAGACCTGGCTGGCCAGCTTGCAACAGCAGCAGCAAGAAGACCAGGTAGCAGAGAAAG AACTTGAGGCATCTGTGCACTTGATGCTTGGCTTCAGACAGATCCTGAGATTCAGCC ATCCCCATTTGTCCAAAGTCAGAAAAGGGTGGTGCACCTGCAGCTCCTGCGGAGACAC ACTCTTCGGGCAGCAGAGCGGAATGTCCGGCGGAAAAGGTCTATTCCAGCTAGAGA GAATCATCAAAAAGCAGAGGCTGCTGGAGGCCAGAAAGAGACTGAGAGAAGCTCAGGAC ATTGTGCTGGCTCCAGGATGACAGCACCCAGGAGCCCCATACCAGGTCTCAGCCCT GATGCCACAGTCCCACGGCTCCATGTAGAAGCAAATTGACGAGTTGCAGTTCTTTGA GCCCCCAAAGACTCTGCAGCAAGCACATGCCCCAGCTACACAGCATTTTCTAAGTTG GGATCCCTCTACCACATTGCCACCTAGGCCTGACCCTACACACCAACATCAGAGAAA ACATCATCAGAAGAGCATTTGCCACAGGCTGCTTCTACCTGCAAGGACAGGGTGCC TCCGCAAGAACGGCTGCATTCTCAGGTGATGGGCAGCCCTGCACAGCCAGAGCAGC CTTGGCCAGGAAGGGAGCCTCAGCTCCAGACGCTTGCTCACCATGAGTCCCAACTCT GTTGGCATCCAGGAAATGGAGATGGGGGTAAAGCAGCCCCATCAGATGGTGAAGCCAGG GCTTAGCATCTCTGAGGAAATCAGCTAACAACTAAAGCCAAGGCATGAGCCAAAGAT CTTACCTCTACTACCAGACCAGAGGGGCGAAGGACTAGCAGACCTAGCCACACA CAAGCTGGGTGGCGAAAAGAAGGGAACCTTGGGACCCACAAGGCTGCTAAGGGAGCCA GTTGCAATTCCTTGTATCTCATGGACCCAGGCAGACTGCTGGGCACGGAAGGCAGT CAAGACTTTTGGACAGAATACAAACACCTTCTCCAAGCAGGGCATCAAAAAGGCAT CAGAGGGTCTGGCAACTAGGGTCAGAAATATTACCAAAAAGTCTCTCACTTGCCCTC TTGGCAGTCCCTTTGAAGAGACAACAAAATACAAGGAGCCAGACACCATGGTCCCCT CACAGATTCAGCCCAGTAATGGATCATTTCAAGAGAAAAGACATGATTTATCTGAC ACAGATAGCAACTACTCATTTGGATTCTCTCATGTGTATGCCAAAGCCCTGATAG	

AGCCACTGAAGCCAGAGGAGAGGAAATGGGATTTCCAGAGCCAGAGAACTCTGAAAG  
 TGATGACAGCCAACTATCTGAGGACTCACTGGCTGAGAAGAGGTACCAAAGCCCCAAA  
 AACAGGCTAGGGGGCAATCGTCCCACCAACAACCGTGGCCAACCCAGGACCAGAACTA  
 GAGCTTCTGTGAGGGGCTTCACTGCAGCCTCAGACAGTGACCTACTTGCTCAAACCTCA  
 TAGGAGCTTCTCCTTGGATAGCCTGATTGATGCAGAGGAAGAACTGGGGGAAGATCAG  
 CAAGAAGAACCTTTCCCTGGTTCAGCTGACGAGATACCCACAGAGACTTTTGGCACC  
 TGGAGGACTCTAGTCTGCCTGTAATGGACCAAGAGGCAATATGCAGGCTTGGTCCCAT  
 CAACTACAGAACAGCAGCTAGGCTGGATGCCGTCCTGCCAATGAGCAGTTCGTTTTAC  
 CTTGATCCTCAGTTCCAACCCCATTTGTGAGCTCCAACCCCATTTGTGAGCTCCAACCCC  
 ATTGTGAGCTCCAGCCCCATTGTGAGCAGGCTGAATCACAGGTAGAGCCAAGCTACTC  
 TGAACAAGCCGACTCTCTCCAAGGCATGCAGCTTTCAAGAGAGAGCCCCACTGATGTCC  
 ATGGATTCTCGGTTTTCTGTGACTCTAAGATCAACCCAGCAGCCCCCAGGAATAG  
 TGGGTTCTTTATGTCCAAGTCCTGATATGCAGGAATTTCACTCCTGTAAGGGGGAGAG  
 GCCTGGATACTGGCCAAATACTGAGGAACTAAAGCCATCAGATGCAGAAACGGTCTTG  
 CCATATAGCTCCAACTGCACCAAGGCAGTACTGAGCTCCTCTGCAGTGCAAGAGATG  
 AGCACACAGCCTCTGCTGCTGATACGCTTAGGCTGTCTCTCTGGGGAATTCAAAGGCT  
 TATTCAACCAGGAGCTGATGGCACCTTTCAGGGCAGATGTATCCTTGACATGACCCAG  
 CAGGGCAGCTCTGAAGCATCCCAATTTCTAGCGTATCAAACGCTGCTGGCTGCCCTGTG  
 CCACCACCTTGACTCATGTAGGCAGCACCCATGAAAGGGATTGGTCTGCCCTTCAGCA  
 GAAGTACCTCCTTGAACCTCTTTGTCTGTTTTGGAGGCCATAGGAGCACCCAAGCCA  
 GCTTACCCCTACCTTGAGGAAGACTCTGGTTCCCTGGCCCAAGCTTCTAGCAAAGGAG  
 GAGATACTCTATTGCCAGTTGGCCCTAGGGTATCTAGCAATCTGAATCTCAACAATTT  
 TCCAGTCCATCTGTCCAGAATCAGGCGTTTGAGGGCAGAGAAAGAACAGGACAGTTTA  
 AATGCCAAATTAGAAGGTGTTTCAGATTTCTTTAGCACTAGTGAGAAAGAGGCGAGTT  
 ATGACGAAACTTATTCCGAGACTTAGAATCATGTCTGCTTCTCGATCTACAATATGC  
 ACAGGTCTTTGCAACAGAGAACGCGATACCAGATTCCATGACAGAAGCATGTGAAGTC  
 AAGCAGAAACAATTTGGAAGAATGCCCTTCAGAGTTGCAGGAAACCTGGACTGATGACTT  
 CCTCTGATGAGGATTTTTCAGAGAAGACGCTTGTCACAGTAATGTCACTACAGCCAC  
 CAAAGCAGACCATTTGGTCCCAAGGCTGGGCTCCTCTCAGGAAAAATAGTGAGTCCAG  
 CCAGGGCAATTAAGTCCCGACAGCCACTACCCACTAGAGGAAGAGAAGACAGATTGCC  
 AGGAGAGCTCTAAGGAAGCAGTTAGAAGACACATAATGTTTCCCTTTGCCCTTCCCTTC  
 AGGTCCAGAGCTATACCTTCACTCTGCTCCTTGAATCCATTGTCTCTTCCCTGTCAG  
 CCCCCACTCTTGAAACATTCTATGTGACCAAAAGCAGGGATGCCCTGACAGAACTG  
 CCTTAGAGATTCCAGCTTGACAGAGAAGTAAGGGTACCTCCCCACCCCCCAGGGAAGC  
 CTGGGGCTTTGGTCAACAACCACCAAGCTCTCCAAGGTGCTTATTTGAAGAATAATTTG  
 CCAGTGCTGTTACAAAACAGAAATTTAAGATTGCCTCATCTCAGCAGGTACAGCTG  
 AGATACCAGTTGATCTGAATACCAGGGAAGTCATCAGAGAATCAGGTAATGCCCTGG  
 AAATATTACAGAAGAAAGCCATGATTGATTGATTCTTCTGTTACTCAGAAACAGCAT  
 TTTCTCCCCTCTACCAGCACAAAAGTATGTGAATTTGAAAACCAAGTTGTAATTTTAA  
 ATAAAAACACAGTTTTCAGCACTTGAGGGAGGAGAGGTCACTGCTCAGTCTGTTG  
 CGGTGCTTCTCAGACAGCACTGAGTCTGGGAAGTCTCTCTCTTTCTGTAATCTGAG  
 GCACGAGAGGAAGAAGAGCTGGATCAGAATACGGTTCTGAGGCAGACCATCAATGTAA  
 GCCTTGAGAAAGACATGCCAGGGGAAAGTGCTGTTTTCTTGAAATCCAGATCAGTAGA  
 TCGTAGAGTAAGCAGCCAGTGATGGTGGCCAGGGTGGTGGCCCAACCCCTAAGTGG  
 GAAGGGAAAAATGAACTGGGCTTCTTGAAAAAGGTCTTCGTCCCAAGATAGCTCAG  
 AAGAGTTTAAGCTTCCAGGTACAAAGCCTGCATATGAAAGGTTCCAGTTAGTTGCATG  
 CCTCAGGAAAGAAACCCAGTGAAATGCAAGTCACAAGAAATGTTAAATCCCAACAGA  
 GAACCTTCTGGAAGAAACAGAATAAAAGAGTTAATAATACTGATGAAATGGCTAGGC  
 TAATTAGGAGTGTAATGCAGCTGGAAAAATGGCATCTTAGAAATTGAATCTAAGCAGAA  
 TAAGCAGGTTCTGCTTCCACACACCAGGAACCGATAAGGAGTTGGTGTTCAGGAC  
 CAGAAGGAGCAGGAGAAGACTGACCATGCCTTTAGGCCAGACAGCTCTGGAAACCCCTT  
 TGCCCTCTAAGGATCAGCCATCTTCTCCAAGACAGACAGATGATCTGTCTTTAGGGA  
 TAGTGAAGCTGGAGCGATGGAGGTTAACAGCATTGGGAACCATCCCCAGGTCCAGAAA  
 ATCCCCCAAAACCCCTTCAGGTCAAGGGAAGGTGTACGAGAGAGTGAACCTGTGAGAG  
 AGCACACACCACTTGACAGAGGTTACCAACACTTCTCTTACCCACAGAGAATGAAA  
 GCATTGGCTAGAGCTCTGCCATTGCAACCCAGGCTAGAGAGGTCTTCTAAGAATAATG  
 GCCAGTTTGTAAAAGCATCAGCAAGTCTCAAAGGGCAGCCTTGGGGCTTAGGAAGTCT  
 TGAGGAATTGGAGACTGTGAAAGGTTTTTCAGGAAAGCCAAGTAGCTGAACACGTAAGT  
 AGTTCCAACCAAGAAGAGCCAAAGCTCAAGGTAAAGTTGAAGAAATGCCATGCAAA  
 GGGGAGGCAGCCTTCAGGAAGAAAATAAAGTGACTCAGAAATTTCTAGTCTCAGCCA

GCTTTGTAGGGACACGTTTTTCAGGCAGGAACTGTGAGCCCATTAAGCCGGACA  
 GAATTCTGTACAGCTCCTCTTCACCAAGACCTGAGTAATACCTTGCCCTTGAATTCTC  
 CAAGGTGGCCAAGAAGGTGTCTTCATGTACCTGTGCTCTAGGCATCTCTTCACTTGA  
 CTGTGTGCTGGATCTCACAATGTTGAAAATTCATAACAGTCCCTTGGTAACTGGAGTA  
 GAGCATCAGGACCAGAGTACGGAGACCAGAAGCCACAGCCCCGAAGGAAATGTTAGAG  
 GGCGTTCCCTCTGAGGCACACACTGCCTGGTGTGGGTCTGTGCGATCCATGGCCATGGG  
 ATCTCATAGTCAATCTGGTGTACCAGAGAGCATTCCTCTGGGGACAGAGGACAGGATC  
 TCAGCAAGCACCAGCCCCAAGACCATGGAAAGGACCTCAGAATCACCTTGCTGGGTT  
 TCAGTACCAGTGAAGATTTTGCTTCTGAAGCCGAGGTGGCTGTACAAAAAGAAATAAG  
 AGTCAGTTCACTGAACAAGGTCTCTAGCCAGCCTGAAAAGAGGGTCAGCTTCTCCTTG  
 GAAGAGGATAGTGACCAAGCCAGCAAGCCAGGAGGAGGAGAGAGATGAGG  
 ACGTCGGACTGACCAAGCGGTGTTTCCCTTAGCACCTGTTTCCCTGCGGAGGGTGCCAG  
 TCCAGAGCCTAGGCTGTGGAGCCCTCTGACCATGCATCCATGTGCTGGCCATCTTG  
 GAGGAGATCAGACAGGCAAAGGCCAGAGAAAGCAGCTTTCATGACTTTGTGGCCAGGG  
 GCACAGTCCCTTTCTTACTGTGAACTTTACTAGAACCCGAATGTTCTTCAAGGGTTGC  
 TGGCAGGCCTCAGTGTAACAAATAGACCAGTCATCATCAGACCAGACCAGGAATGAG  
 GGTGAAGCACCGGATTTTCATGTGGCATCTCTATCTGCTGAAGCAGGGCAGATAGATC  
 TGTACCTGATGAGAGGAAAGTCCAGGCCACATCTCTGTCTGCAGACAGCTTTGAATC  
 TCTGCCCAATACGGAAGTACAGAGAGCCATGGGATCCTGTGCAGGCTTTCTCCCAT  
 GCTGCTCCTGCTCAAGACAGGAAACGTCGTACTGGAGAACTGAGGCAGTTCCGCGGGAG  
 CAAGTGAACCATTTATATGTCACTCTAGTTCCTTCTGAAATCATAGAGAAAAAGAAAGA  
 TGCAACCAGAACACCTTCTCAGCTGATCCTTTGGCCCCAGACAGTCCCTCGTTCTTCA  
 GCACCTGTGGAGGAGGTGAGGAGGTAGTATCAAAGAAGGTAGTGGCTGCCTTACCTT  
 CTCAGGCCCTTATGATGATCCTAGAGTGACTCTGCATGAGCTAAGTCAGTCAGTTCC  
 GCAGGAGACTGCAGAGGGCATAACCCCTGGCAGTCAGGACAGCAGCCCAGAGCATCAG  
 GAACCCAGAACTCTAGACACCACATATGGAGAA&TTTTAGATAATTTGTTAGTGACTG  
 CACAGGGAGAAAAAACAGCCCATTTTGAAGTCAGTCTGTGACCTGTGATGTTTCAGAA  
 TTCTACAAGTGCCCTCAGGGCCTAAGCAAGACCATGTCCAATGCCCTGAGGCTTCTACT  
 GGCTTTGAAGAAGGTAGGGCAAGTCCCAAACAAGATACCATTTCTGCCTGGAGCTCTGA  
 CAAGGGTTGCACTGGAAGCTCCACACAGCAGTGTGTGCAGTGTAAGGAGAGTGTGG  
 GTCTGGGTTGACAGAAGTCTGCAGGGCTGGCAGCAAACATTCAGGCCAATTCCACTG  
 CCAGATCAAAGACCAAGCGCAAATCCTGGGGGAATTGGGGAGGAAGCCCCATGTAGAC  
 ACCCAAGGGAAGCTTATAGATGGCCCTGTCTTCTCAAGGAACCTGAAGGCAGCAGGAC  
 TCTCAGCCCGTCTAGAGGGAAGAGAGCAGAACTCTTCCCTGCGGACAGCCATGCAGT  
 TCTCAACCTGTTGCTACTCATGCTTATTCCTCCCATTCCTCTACTTTACTGTGTTTAA  
 GAGATGGTGACCTAGGGAAGGAGCCTTTCAAGGCTGCCCCACATACTATCCACCCACC  
 CTGTGTAGTACCTTCCAGGGCCTATGAAATGGATGAGACAGGAGAGATCTCTAGGGGA  
 CCTGATGTGCACTTGACACATGGCCTTGAGCCCAAAGATGTTAACAGGGAATTTAGGC  
 TAACAGAGAGCAGCACTTGTGAGCCTTCTACTGTGGCTGCTGTCTTCTCGAGCTCA  
 AGGCTGCAGATCCCCTTCTGCTCCTGACGTGAGGACAGGTTCTTCAGCCACTCAGCT  
 ACTGATGGAAGCGTGGGGTTAATAGGGGTTCTTGAGAAAAAGGTTGCTGAGAAGCAAG  
 CAAGCACAGAACTTGAGGCTGCCTCTTCCCTGCAGGCATGTACTCTGAGCCCCTGAG  
 GCAGTTTAGGACAGCTCTGTAGGTGACCAGAATGCACAGGTGTGTCAAACCAATCCA  
 GAACCACTGCAACAACCTCAGGGACCACACACCTGGATTTAAGTGAAGGGTCTGCTG  
 AGAGCAAGTTGGTGGTAGAGCCACAGCATGAATGTTTAGAAAATACCACTAGATGTTT  
 TTTGAAAAGCCACAATTTTCCACTGAGTTGAGGGATCACAATCGCTTGGATTCCCAA  
 GCCAAGTTTGTAGCAAGGTTAAAAACATACCTGCAGCCCCCAGGAAGACAGTCCCTGGC  
 AGGAAGAAGAGCAGCACAGAGACCAGGCTTCAGGTGGTGGGAAGGCTTCGCCCAGGG  
 TGTGAATCCCCCTTCTGATGAAGATGGCTTAGATGGCTGTGAGATTTAGATGCTGGG  
 AGAGAGGAGGTGGCTGTGGCCAAGCCTCCTGTGTCCAAGATTTTATCACAGGGCTTCA  
 AAGACCCAGCCACTGTGTCTTGAGGCAAAATGAAACACCGCAGCCTGCTGCTCAGAG  
 GAGTGGCCACCTTACACTGCGCAGAGAGCAGCCAGCACCCAACCAAGGGGCTCACTT  
 CCTGTGACTACAATCTTCTTGGCCCCAAACACTCCAGGTCTCCCCACACCACAGT  
 TCTCAGTTGTGCGCTCTTCTCGTTCTTTCAGGAGCTGAACCTGAGTGTGGAGCCTCC  
 TTCCCTACAGACGAAGATACACAGGGGCTAACAGATTGTGGAACCCACATCTCAGG  
 GGCTATTCTCAGGAAAGTCAGTGGCAAGAACATCTCTGCAGGCTGAGGACAGCGATC  
 AGAAAGCCTCATCTCGCTTGGATGATGGGACTACCGATCACAGGCACCTGAAGCCTGC  
 CACCCCTCCTTATCCAATGCCTTCCACTCTCTCACACATGCCAACCCCTGATTTACG  
 ACCAGCTGGATGTCTGGTACTTTGGAACAAGCCCAACAGGGAAAGCGAGAGAACTGG  
 GTGTCCAGGTTAGGCCAGAAAAATTGGTGCTCTCAGATGGACAAAGGAATGCTGCACTT  
 TGGCTCCAGTGACATCAGTCCCTATGCGCTGCCGTGGCGTCCGGAGGAGCCTGCACGT

ATCAGCTGGAAGCAGTATATGTCTGGCAGTGCAGTCGATGTTTCCTGCAGCCAGAAGC  
 CCCAGGGGCTGACACTATCAAATGTGGCCCGGTGCTCCAGCATGGACAATGGCCTAGA  
 AGACCAGAACTCCCCTTTCCACTCCCACCTCAGCACTTACGCCAATATTTGTGATCTG  
 TCAACCACACACAGCAGCACTGAGAATGCCAGGGTTCAAATGAGGCCTGGGAAGTAT  
 TCCGAGGGAGTTCTTCAATTGCCTTAGGAGACCCCCACATCCCGACGAGCCCTGAAGG  
 AGTAGCCCCCACTTCGGGTCTATGACAGAAGGCCTCAGTTCAGGGGCCCTTCTGGTGAA  
 GCAGACTGTCTGAGGAGTAAGCCCCCTTGGCCAAAGGAAGTGCTGCAGGTCAGTG  
 ATGAGATTATGCTGCTGTATCCATCAGAGGCAGGCTGCCCTGTGGGACAGACCAGGAC  
 GAACACATTTCGAACAGGGCACACAGACCCCTCGGCAGCAGGCGCCACTGGAGCAGCACT  
 GACATCTCCTTTGCTCAGCCTGAAGCCAGTGCACTATCAGCCTTTGATCTGGCCTCAT  
 GGACCAGCATGCACAATCTGTCTCTCCACCTCTCACAGCTCCTGCACAGTACCTCAGA  
 GCTGCTTGGGAGTCTCTCCAGCCAGATGTGGCCAGAAGGGAGCAGAACACCAAGAGG  
 GACATCCCAGATAAAGCCCCACAGGCCCTGATGATGGATGGCTCTACTCAGACCACTG  
 TGGATGAGGGCAGCCAGACTGACCTCACCTTACCCACCCTGTGCCTCCAGACTTCAGA  
 GGCTGAACCTCAGGGAGCCAAATGTGATCCTTGAAGGGCTAGGCTCAGATACTCCTGACT  
 GTGTCTCAAGAAGAGGGAGATGTGCCAGGGGTACCTCAGAAGAGAGAGGCAGAGGAAA  
 CAGCACAGAAAATGGCTCAGCTCCTCTATCTTCAGGAAGAAAGCACTCCCTACAGCC  
 CCAGAGCCCTTCAATACCTCATCCCACTTGAGGTTTCAGAAAGCCCCGTTGGGCAG  
 CATCTTCCTTCTGTGAGCCCTCAGTTTCTGATGCTTTCTGCCTCCAGCTCCCAGC  
 CAGAGGAGTCATATGCTTAGTTGTGTCAGCAGTCCCAGTCCCAGCTCCCCTCATTCCCC  
 AGGGCTCTTTCCAGTACTTCCGAGTATCCTGGGGACTCCAGGGTCCAGAAGAAGCTG  
 GGCCCCACAAGTGCTTTGTTCGTGGACAGGGCTCCTCCCCAATCCTCACTCTTAGTG  
 CCAGCACCCAAAGACCGGGTCTTTCCCGAGGCTCTTTGACCTCTCAGCCCCCTCAAC  
 TCACCTGTGTAAGGCCACCAGAAGCTTGACTCCAGCCCAGACCTGTTGATGCCCCA  
 AGGACTCCAATGGATAATTATTCCAAACCACTGACGAGTTAGGTGGCTCCAGAGAG  
 GTAGAAGTTCCTTACAAAGGAGTAATGGGAGATCCTTCTTGAAGTGCACCTCCCCACA  
 CAGCCACAGCAGAGTCCAAACTCCAATTTAGTTTCTTAGGGCAGCACCTCAGCAG  
 CTTTCAGCCCAGGACAACATATCGGGGTCCAAAGCAGACTGCTGCCACCACCACTGAGGC  
 ACAGGAGCCAAAGGCTGGGCAACAGCTTTGTGCTGAGAAGGTGGCTTCCCCGGAGCA  
 TTGCCCACTGAGCGGTAGGGAGCCAAGTCAGTGGCAGAGCAGGACAGAAAATGGAGGT  
 GAGAGTTCAGCATCTCCAGGGGAACCAACGCACTCTGGACCGACCTTCTTCATGGG  
 GAGGCTCCAGCACCTCAGCCCCCTGCCCTGTCTCTGAGTTGACTGATACTGCAGGGCT  
 CCGAGTTCTGCCTTGGGCTTCCCTCAGGCTGCCAACCCTGAGGAGTTACTGTGCTTC  
 AGTTGCCAGATGTGCATGGCCCCCTGAGCACCAGCACCACAGTCTGAGGGACCTCCCGG  
 TGCATAACAAATTTAGTAAGTGGTGTGGGGTTTCAAGGGCTCACCTGGGGGTGGA  
 CATGACTGAGGAGGAGCTGGGGCCAGCGGTGATCTCAGCTCTGAAAAGCAGGAACAG  
 AGTCCCCCACAACTCCTAATGACCACAGCCAGGATTTGAGTGGTCCAAGAGGGAGC  
 AGATCCCCCTGCAAGTTGGGGCCAGAACCTCTCACTCAGCTGGAACTCAGAGAAGC  
 GAAACTGCACCATGGCTTTGGGGAGGCCGATGCCCTGCTCCAGGTGCTGCAGAGTGGG  
 ACAGGGGAGGCGCTTGTGCTGATGAACCTGTGACATCCACCTGGAAGGAGCTCTATG  
 CACGGCAAAAAAGGCCATTGAGACCTCAGGAGAGAGCGGGCTGAGCGACTTGGGAA  
 CTTCTGCCGGACGCGAAGCCTTAGCCCTCAGAAACAACTGAGCCTCCTGCCCAACAAA  
 GATCTCTTCATCTGGGATCTGACTTGCCAGCAGACGCGGAGAATACCTGCAGCAAC  
 TGAGGAAGGATGTTGTGGAGACCACAGGAGCCAGAGTCAGTGTCAAGGTGAGTCA  
 CACACCCTCTGACATAGAGTTGATGCTGCAAGACTACCAGCAGGCCCATGAGGAGGCC  
 AAGGTGGAGATTGCCCGGGCCGAGACCAACTGCGGGAGCGGACTGAACAAGAGAAGC  
 TGAGAATCCACCAGAAGATCATTTCCAGCTATTGAAGGAAGAGGATAAACTACATAC  
 CTTGGCCAATTCCAGCTCCCTGTGCACCAGCTCTAATGGAAGCCTCTCGTCTGGCATG  
 ACCTCTGGCTATAATAGCAGCCCCAGCCTTGTTCAGGCCAGCTCCAGTTCCAGAGAATA  
 TGGGGCATAACAACTTGCCCTGATTCCAGGGATGTATGGATAGGGGATGAGCGAGGAGG  
 CCATTCTGCAGTGAGGAAGAACTCTGCCTACAGCCACAGAGCCTCCCTGGGCAGTTGC  
 TGCTGTTTACCATCCAGTCTGTCCAGCTTGGGGACCTGCTTTTCTCCTCCTACCAGG  
 ATTTGGCCAAGCATGTCGTGGACACTTCTATGGCTGATGTAATGGCTGCTTGTTCGGA  
 TAATTTGCACAACCTCTTCAGCTGCCAGGCAACTGCTGGCTGGAACATCAGGGTGAG  
 GAGCAGGCGGTGCAGCTTTACTACAAGGTGTTTTCTCCCACTCGCATGGCTTCTGG  
 GGGCAGGTGTGGTGTCCAGCCGCTGTCTCGTGTGTGGGCGGCTGTCACTGACCCAC  
 TGTGTGGCCCCGTATTACAAGCCCATCCAGACAGCAAGGCTGCATCAGCGAGTGACC  
 AACAGCATCAGCCTGGTGTACTTGGTGTGCAACACCACCCTGTGCGCACTGAAGCAGC  
 CACGGGATTTCTGTTGTGTCTGCGTGGAGCCAAAGAGGGTCACTGTCTGTCTATGGC  
 AGCCAGTCTGTGTATGATACATCCATGCCAAGACCAGCAGAAAAATGGTTACCGGG  
 GAGATCCTGCCAGTGCCTGGATCTTGCAAGCCATCACTGTGGAAGGAAGGAAGTCA

	CCAGAGTCATCTACTTGGCCAGGTGGAACCTGGTGCTCCAGGCTTCCACCTCAGCT CCTGAGCTCTTTTCATCAAACGGCAGCCACTGGTTATAGCCAGACTGGCTTCTTCTCTT GTGCAGGAAAAGCTGATGCTACCTGCTGTGGCCGATTGGGGCAGACAGCACTGGCCCA GGGATGCTAGCAAAGCCAGTCAGTACTTGGTCACAGCTGGCACCAGTGCAGAGCAAA CGGCCTGAGCTCCTGGCCAGACTATCCAGAGTGAATGCAGCTCTGCTCACCTTTTGG ATTTCTCACCTTTCTTCTGTTTCTGGGACTCTGCGGCAGACAGGACACTTAAGGAC CAGGACTGGGCACAGCCAGCAGAGCCGGGGACTGCAGTGCCTTGGCAAGGTGCTTCCG CAGGCTGGTAGGGAA		
	ORF Start: ATG at 1		ORF Stop: TAA at 14320
	SEQ ID NO: 322	4773 aa	MW at 524614.9kD
NOV34a, CG157505-01. Protein Sequence	MANVQVAVRVRPLSKRETKEGGRIIVEVDGKVAKIRNLKVDNRPDGFGDSREKVMFAG FDYCYWSVNPEDPQYASQDVVFQDLGMEVLSGVAKGYNICLFAYGQTGSGKTYTMLGT PASVGLTPRICEGLFVREKDCASLPSSCRIKVSFLEIYNERVRDLLKQSGQKKSYYTLR VREHPMGPPYVQGLSQHVVTNYKQVIQLLEEGIANRITAATHVHEASSRSHAIFTIHY TQAI LENNLPSEMAKINLVDLAGSERADPSYCKDRIAEGANINKSLVTLGIVISTLA QNSQVFSQCQLNLSVSNNGGDSGILSSPSGTSSGGAPSRQSYIIPYRDSVLTWLLKDS LGGNSKTIMVASVSPAHTSYSETMSTLRYASSAKNIINKPRVNEDANLKLIRELREEI ERLKALLSFELRNFSLSDENLKEVLQNELKIDQLTKDWTQKWNWQALMEHYSVD INRRRAGVVIDSSLPHLMALEDDVLSTGVVLYHLKEGTTKIGRIDSDQEQDIVLQGGW IERDHCTITSACGVVLRPARGARCTVNGREVTASCLTQGA VITLGAQKFRFNHPA EAAVLRQRQVGEAAAGRGSLEWLDLDGDLAASRLGLSPLLWKERRALEEQCEDHDQT PRDGETSHRAQIQQQSYVEDLRHQILAEEIRA AKELEFDQAWISQIKENQQLLRE ETWLASLQQQQQEDQVAEKELEASVALDAWLQTDPEIQPSFVQSKRVVHLQLLRH TLRAAERNVRKKVSFQLERI IKQRLLLEAQKRLKLTTLCLWLQDDSTQEPYQVLS DATVPRPPCRSKLTSCSSLSQRLCSKHPQLHSIFLSWDPSTTLPPRPDPHTQTSK TSSEEHLPQAASYPARTGCLRKNGLHSSGHGQPCTARAALARKGASAPDACTMTSPNS VGIQEMEMGVKQPHQMV SQGLASLRKSANKLKPRHEPKIFTSTTQTRGAKGLADPSHT QAGWRKEGNLGTAKAGASCNSLYPHGPRQTAGHGKAVKTFWTEYKPPSPSRASKRH QRVLATRVRNITKSSHLPLGSPKLRQONTRDPTMVP LTFSPVMDHSREKDNLDSD TDSNYSLSLSCVYAKALIEPLKPEERKWDFFEPENSESDSLSLSESLAEKRYQSPK NRLGGRPTNNRGQPRTRTRASVRGFTAASDSDLLAQTHRSFSLDSLIDAEELGEDQ QEEPPFGSADEIPTETFWHLEDSSLPVMDQEAICRLGPINRYTAARLDAVLPMSSSFY LDPQFQPHCELQPHCELQPHCEQAESQVEPSYSEQADSLQGMQLSRESPLMS MDSWFSCDSKINPSPPGIVGSLCPSDPMQEFHSCKGERPGYWPNTTELKPSDAETVL PYSSKLHQGSTELLCSARDEHTASAADTSRLSLWGIQRLIQPGADGTFQGRCI PDMTQ QGSSEASHNSSVSNVLAASATTLTHVGSTHERDWSALQQKYLLELSCPVEAIGAPKP AYPYLEEDSGSLAQASSKGGDTLLPVGPRVSSNLNLNFPVHLSRIIRLRAEKEQDSL NAKLEGVSDFFSTSEKEASYDETY SADLESLSASRSTNAQVFATENAI PDSMTEACEV KQNNLEELQSCRKPLGKMTSDEDFQKNACHSNVTTATKADHNSQGWAPLRKNSAVQ PGQLSPDSHYPLEEEKTDCQESSKEAVRRHINVSFALPSGPELYLHSAWPNPLSSSLQ PPLLETIFYVTKSRDAL TETALEI PACREVRVPSPPPREAWFGHNHQA LQGAYLKNNL PVLLQNQNSKIASSQVTA EI PVDLNTREVIRESGKCPGNITEESHDSVYSSVTQNRH FLPSTSTKVCFENQVVILNKKHSFPALLEGGEVTAQSCCGASSDSTESGKSLLFRESE AREEEELDQNTVLRQTINVSLEKDMPGESAVSLKRSVDRRVSSPVMVAQGGGPTPKW EGKNETGLLEKGLRPKDSSEEFKLPGTKPAYERFQLVACPQERNPSECKSQEMLNPNR EPSGKKQNKRVNNTDEMARLIRSVMLENGILEIESKQNKQVHASHTPGTDKELVFQD QKEQEKTDHAFRDPSSGNPLPSKDQPS PRQTD DTVFRDSEAGAMEVNSIGNHPQVQK ITPNPFRSREGVRESEPVREHHPAGSDRPARDICDSL GKHTTCREFTNTSLHPQRMK ALARALP LQPRLES SKNNGQFVKASASLKGQPWGLGSLELETVKGFOESQVAEHVS SSNQEEPKAQGVVEEMPMQRGGS LQEENKVTQKFP SLSQLCRDTFFRQVTVSPLLSRT EFTAPLHQDLSNTLPLNSPRWPRRCLHVPVALGISSLD CVLDLTLMLKIHNSPLVTGV EHQDQSTETRSHSPEGNVGRSSEAH TAWCGSVRSMAMGSHS QSGVPES IPLGTEDRI SASTSPQDHGKDLRITLLGFSTSEDFASEAEVAVQKEIRVSSLNKVSSQPEKRVFSL EEDSDQASKPRQKAEKETEDVLTSGVSLAPVSLPRVPSPEPRLLPSDHASMLAIL EEIRQAKAQRQLHDFVARGTVLSYCTILLEPECSSRVAGRPOCKQIDQSSSDQTRNE GEAPGFHVASLSAEAGQIDLLPDERKVQATSLSADS FESLPNTETDREPDPVQAFSH AAPAQDRKRTGELRQFAGASEPFICHSSSEIIEKKKDATRTPSSADPLAPDSPRSS APVEEVRRVSKKVVAALPSQAPYDDPRVTLHELSSQSVPOETAEGIPPGSQDSSPEHQ EPRTLDTTYGEVSDNLLVTAQGEKTAHFESQSVTC DVQNSTASAGPKQDHVQCPEAST		

<p>GFEEGRASPKQDTILPGALTRVALEAPTQQCVQCKESVGSGLTEVCRAKSHSRPIPL  PDQRPSANPGGIGEEAPCRHPREALDGPVFSRNPEGSRTLSPSRGKESRTLPCRQPCS  SQPVATHAYSSHSSTLLCFRDGDLGKEPFKAAPHTIHPPCVVPSRAYEMDETGEISRG  PDVHLTHGLEPKDVRNREFRLTESSTCEPSTVA AVL SRAQGC RSPSAPDVRTGSF SSHA  TDG SVGLIGVPEKKVAEQASTELEAASFAGMYSEPLRQFRDSSVGDQNAQVCQTNP  EPPATTQGPHTLDLSEGSAESKLVEPQHECLENTRCFLEKPKQFSTELRDHNRDLSQ  AKFVARLKHTCSPQEDSPWQEEEQHRDQASGGGEGFAQGVNPLPDEDGLDGCQILDAG  REEVAVAKPPVSKILSQGFKDPATVSLRQNETPQPAQQRSGHLYTGREQPAPNHRGSL  PVTTFSGPKHSRSSPTPQFSVVGSSRLQELNLSVEPPSPPTDEDTQGGPNRLWNPHLR  GYSSGKSVARTSLQAEDSDQKASSRLDDGTTDHRHLKPATPPYPMPSTLSHMPPTDFT  TSWMSGTLEQAQQGKREKLGQVRPENWCSQMDKGMLHFGSSDISPYALPWRPEEPAR  ISWKQYMSGSAVDVSCSQKPQGLTSLNVARCSSMDNGLEDQNSPFHSHLSTYANICDL  STTHSSTENAQGSNEAWEVFRGSSSIALGDPHIPTSPGVAPTS GHDRRPQFRGPSGE  ADCLRSKPPLAKGSAAGPVDEIMLLYPSEAGCPVGQTRTNTFEQGTQTLGSRHRWSST  DISFAQPEASAVSAFDLASWTS MNHLSLHLSQLLHSTSELLGSLSQPDVARREQNTR  DIPDKAPQALMMDGSTQTTVDEGSQTDLTPLTCLQTSEAE PQANVILEGLGSDTST  VSQEEGDVPGVPQKREAEETAQKMAQLLYLQEESTPYKQSPSIPSSHRLRFQKAPVGQ  HLPSPSPSVDAFLPPSSQPEESYCLVSSPSPSSPHSPGLFPSTSEYPGDSRVQKKL  GPTSALFVDRASSPILTLASTQEPGLSPGSLTSLAPSTHPVEGHQKLDSSPDVPDAP  RTPMDNYSQTTDELGGSQRRSSLQRSNGRSFLELHSPHSPQSPKLFQSFGLQHPQQ  LQPRTTIGVQSRLPLPPLRHRSQRLGNSFVPEKVASPEHCPLSGREPSQWQSRTEGG  ESSASPGEPQRTLDRPSSWGGLOHLSPCPVSELTDTAGLRGSALGLPQACQPEELLCF  SCQMCMAPBHQHSLRDLPVHNKFSNWCQVQKQSGPGLDMTEELGASGDLSSSEKQEQ  SPPQPPNDHSDSEWSKREQIPLQVGAQNLSLSVELTEAKLHHGFGEADALLQVLQSG  TGEALAADEPVTSTWKELYARQKKAIE TLRRERAERLGNFCRTRLSLPQKQLSLLPNK  DLFIWDLPLSRREYLQQLRKDVVETTRSPESVSRSAHTPSDIELMLQDYQQAHEEA  KVEIARARDQLRERTEQEKLRIRHQKII SOLLKEEDKLHTLANSSSLCTSSNGSLSSGM  TSGYNSSPALSGQLQFPENMGHTNLPDSRDVWIGDERGGHSAVRKNSAYSHRASLGSC  CCSPSSLSSLGTCFSSSYQDLAKHVVDTS MADVMAACSDNLHNLFSQATAGWNYQGE  EQAVQLYYKVFSPTRHGFLGAGVVSQPLSRVWAAVSDPTVWPLYKPIQTARLHQRT  NSISLVYLVNCTTLCAKQPRDFCCVCVEAKEGHLSVMAAQSVYDTSMPRPSRKMVHG  EILPSAWILQPTVEGKEVTRVIYLAQVELGAPGFPQLLSSFIKQPLVIARLASFL  VQEKMLMPAVADWGRQHWPRDASKAQSVLGHSHWQCRANGLSSWPRLSRVNAALLTFW  ISHLSFLFLGLCGRQDT</p>
---

Further analysis of the NOV34a protein yielded the following properties shown in Table 34B.

Table 34B. Protein Sequence Properties NOV34a	
PSort analysis:	0.9000 probability located in nucleus; 0.6640 probability located in plasma membrane; 0.3694 probability located in mitochondrial inner membrane; 0.3000 probability located in microbody (peroxisome)
SignalP analysis:	No Known Signal Sequence Predicted

5. A search of the NOV34a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 34C.

Table 34C. Geneseq Results for NOV34a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV34a Residues/ Match	Identities/ Similarities for the Matched	Expect Value

		Residues	Region	
AAU74557	Human kinesin motor protein HsKif16a - Homo sapiens, 563 aa. [US6333184-B1, 25-DEC-2001]	1..590 1..563	518/591 (87%) 519/591 (87%)	0.0
AAU74558	Human kinesin motor protein HsKif16a motor domain - Homo sapiens, 357 aa. [US6333184-B1, 25-DEC-2001]	1..385 1..357	334/385 (86%) 335/385 (86%)	0.0
ABB61704	Drosophila melanogaster polypeptide SEQ ID NO 11904 - Drosophila melanogaster, 1174 aa. [WO200171042-A2, 27-SEP-2001]	23..784 4..707	306/782 (39%) 439/782 (56%)	e-132
AAM40034	Human polypeptide SEQ ID NO 3179 - Homo sapiens, 893 aa. [WO200153312-A1, 26-JUL-2001]	2..737 4..763	295/804 (36%) 416/804 (51%)	e-117
ABP51294	Human MDDT SEQ ID NO 316 - Homo sapiens, 757 aa. [WO200240715-A2, 23-MAY- 2002]	2..609 19..591	248/619 (40%) 355/619 (57%)	e-114

In a BLAST search of public sequence databases, the NOV34a protein was found to have homology to the proteins shown in the BLASTP data in Table 34D.

Table 34D. Public BLASTP Results for NOV34a				
Protein Accession Number	Protein/Organism/Length	NOV34a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q9P2P6	KIAA1300 protein - Homo sapiens (Human), 1820 aa (fragment).	2881..4698 1..1818	1818/1818 (100%) 1818/1818 (100%)	0.0
Q9H6S2	CDNA: FLJ21936 fis, clone HEP04408 - Homo sapiens (Human), 818 aa (fragment).	1080..1883 1..804	802/804 (99%) 802/804 (99%)	0.0
Q9DDA6	Kinesin-like protein - Xenopus laevis (African clawed frog), 1499 aa (fragment).	1..1285 1..1269	617/1321 (46%) 825/1321 (61%)	0.0
Q15885	Partial cDNA sequence, clone x529, unknown open reading frame - Homo sapiens (Human), 380 aa (fragment).	1428..1807 1..380	378/380 (99%) 378/380 (99%)	0.0
AAH32885	Hypothetical protein - Mus musculus (Mouse). 371 aa	4340..4698 1..369	284/370 (76%) 315/370 (84%)	e-158

	(fragment).			
--	-------------	--	--	--

PFam analysis predicts that the NOV34a protein contains the domains shown in the Table 34E.

Table 34E. Domain Analysis of NOV34a			
Pfam Domain	NOV34a Match Region	Identities/ Similarities for the Matched Region	Expect Value
kinesin	9..295	122/340 (36%) 219/340 (64%)	3.1e-85
kinesin	332..413	52/83 (63%) 72/83 (87%)	7e-41
FHA	503..569	24/80 (30%) 46/80 (58%)	0.0059
REV	4268..4335	16/69 (23%) 43/69 (62%)	0.52
START	4496..4704	45/254 (18%) 138/254 (54%)	0.012

### Example 35.

5 The NOV35 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 35A.

Table 35A. NOV35 Sequence Analysis			
	SEQ ID NO: 323	2039 bp	
NOV35a, CG157629-01 DNA Sequence	CTAAGAGTGGTTCTCGCAGCTTAAAGGGAGGCACTTTTCACACTCTGTCTTAAATC AGAAGTTGAATTCATGAACACATATGATTAGATAGAAGTCATGGGATGCAGCAGTTC TTCACGAAAACAGGAGATCTGACACATCACTGAGAGCTGCGTTGATCATCCAGAAC TGGTACCGAGGTTACAAAGCTCGACTGAAGGCCAGACAACACTATGCCCTCACCATCT TCCAGTCCATCGAATATGCTGATGAACAAGGCCAAATGCAGTTATCCACCTTCTTTTC CTTCATGTTGGAAAACACACACATATACATAAGGAAGAGCTAGAATTAAGAAATCAG TCTCTTGAAAGCGAACAGGACATGAGGGATAGATGGGATTATGTGGACTCGATAGATG TCCCAGACTCCTATAATGGTCTCGGCTACAATTTCTCTCACTTGTTACGGATATTGA TTTACTTCTTGAGGCCTTCAAGGAACAACAGATACTTCATGCCCATTTATGTCTTAGAG GTGCTATTTGAAACCAAGAAAGTCTGAAGCAAATGCCGAATTTCACTCACATACAAA CTTCTCCCTCCAAAGAGGTAACAATCTGTGGTGATTTCATGGGAACTGGATGATCT TTTTTTGATCTTCTACAAGAATGGTCTCCCTCAGAGAGGAACCCGATGTTTTTAAT GGTGACTTTGTAGATCGAGGAAGAATTCCATAGAGATCCTAATGATCCTGTGTGTGA GTTTCTTGCTTACCCCAATGACCTGCACTTGAACAGAGGGAACCCGAGATTTTAT GATGAATCTGAGGTATGGCTTACGAAAGAAATTTGCATAAATATAAGCTACATGGA AAAAGAATCTTACAAATCTTGAAGAATTCTATGCCCTGGCTCCCAACGGAAACAAACA GAGACCATGGCACTGACTCGAAGCACAATAAAGTAGGTGTGACTTTTAAATGCACATGG AAGAATCAAAACAAATGGATCTCCTACTGAACACTTAACAGAGCATGAATGGGAACAG ATTATTGATATTCTGTGGAGTGATCCCAGAGGCAAAAATGGCTGTTTCCAAATACGT GCCGAGGAGGGGCTGCTATTTTGGACCAGATGTTACTTCCAAGATTCTTAATAAATA CCAGTTGAAGATGCTCATCAGGTCTCATGAATGTAAGCCGAAGGGTATGAAATCTGT		

	CATGATGGGAAGGTGGTGACTATATTTCTGCTTCTAATTATTATGAAGAAGGCAGCA ATCGAGGAGCTTACATCAAACATATGTTCTGGTACAACCTCCTCGATTTTTCCAGTACCA AGTAACATAAGCAACGTGCTTTTCAGCCTCTTCGCCAAAGAGTGGATACTATGGAAC AGCGCCATCAAGATATTAAGAGAGAGAGTGATTTCACGAAAAAGTGACCTTACTCGTG CTTTCCAACCTTCAAGACCACAGAAAAATCAGGAAAACTTTCTGTGAGCCAGTGGGCTTT TTGCATGGAGAACATTTTGGGGCTGAACTTACCATGGAGATCCCTCAGTTCGAATCTG GTAAACATAGACCAAAATGGAACGTTGAATACATGTCCAGCTTCCAGAATATCCGCA TTGAAAAACCTGTACAAGAGGCTCATTCTACTCTAGTTGAACTCTGTACAGATACAG ATCTGACCTGGAATCATATTTAATGCCATTGACACTGATCACTCAGGCCTGATCTCC GTGGAAGAATTTCTGTCCATGTGGAACCTTTTCTAGTTCTCACTACAATGTTTCACTTG ATGATTTCCCAAGTCAATAAGCTTGCCAACATAATGGACTTGAAACAAAGATGGAAGCAT TGACTTTAATGAGTTTAAAGGCTTTCTATGTAGTGATGATAGATATGAAGACTTGATG AAACCTGATGTACCAACCTTGGCTAAACACAAATGAGAGCTTCCCTCAGGCTCCCTG AAACAGCTAGGCCAAATCACAAGTACAGTCTTCCAACACCCCTGAAATTCATAGT CAGTAGCAG		
	ORF Start: ATG at 100		ORF Stop: TAA at 1939
	SEQ ID NO: 324	613 aa	MW at 71315.2kD
NOV35a, CG157629-01 Protein Sequence	MGCSSSTKTRRSDSLRAALIIQNWYRGYKARLKARQHYALTIFQSIEYADEQGMQ LSTFFSFMLENYTHIHKEELELRNQSLSEQDMRDWDYVDSIDVPDSYNGPRLQFPL TCTDIDLLEAFKEQILHAHYVLEVLFTKVKVQKMPNFTHIQTSPEKVTICGLDH GKLDLFLIFKNGLPSEPNPYVFNDFVDRGKNSIEILMLLCVSFLVYPNDLHLNRG NHEDFMMNLRYGFTKEILHKYKLHGKRIQLIEEFYAWLPETNRDHGTDKSKHNKVG TFNAHGRIKNGSPTEHLTEHEWEQIIDILWSDPRGKNGCFNCRGGGCGYFGPDVTS KILNKYQLKMLIRSHECKPEGYEICHGKVVITFSASNYEEGSRNGAYIKLCSGTP RFFQYQVTKATCFQPLRQVDTMENSARKILRERVSRKSDLTRAFQLQDHRKSGKLS VSQWAFCMENILGLNLPWRSLSNLVNIQNGNVEYMSFQNIKIEKPQEAHSTLVE TLYRYSDELIIFNAIDTDHSLISVEEFAMWKLFSHYNVHIDDSQVNKLANIMDL NKDGSIDFNEFLKAFYVVHRYEDLMKPDVTNLG		
	SEQ ID NO: 325	2039 bp	
NOV35b, CG157629-01 DNA Sequence	CTAAGAGTGGTTCCTCGCAGCTAAAGGGAGGCACTTTTCACACTCTGTCTAAAATC AGAAGTTGAATTCATGAACACATATGATTTAGATAGAAGTCAGGATGCAGCAGTTC TTCAACGAAAACAGGAGATCTGACACATCACTGAGAGCTCGGTTGATCATCCGAAC TGGTACCGAGGTTACAAAGCTCGACTGAAGGCCAGACAACACTATGCCCTCACCATCT TCCAGTCCATCGAATATGCTGATGAACAAGGCCAAATGCAGTTATCCACCTTCTTTTC CTTCTATGTTGGAAACTACACACATATACATAAGGAAGAGCTAGAATTAAGAAATCAG TCTCTTGAAAGCGAACAGGACATGAGGGATAGATGGGATTATGTGGACTCGATAGATG TCCCAGACTCCTATAATGGTCTCGGCTACAATTTCTCTCACTGATCCGATATTGA TTTACTTCTTGAGGCCTTCAAGGAACAACAGATACTTCATGCCCATATGTCTTAGAG GTGCTATTTGAAACCAAGAAAGTCTGAAGCAAATGCCGAATTTCACTCACATACAAA CTTCTCCCTCCAAAGAGGTAACAATCTGTGGTGATTTGCATGGGAACTGGATGATCT TTTTTTGATCTTCTACAAGAATGGTCTCCCTCAGAGAGGAACCCGTATGTTTTTAAT GGTGACTTTGTAGATCGAGGAAAGAATTCATAGAGATCCTAATGATCCTGTGTGTA GTTTTCTGTCTACCCCAATGACCTGCACTTGAACAGAGGGAACCCAGAGATTTTAT GATGAATCTGAGGTATGGCTTCACGAAAGAAATTTGCATAAATAAAGCTACATGGA AAAAGAATCTTACAAATCTTGAAGAATTCTATGCCTGGTCCCAACGGAACAAACA GAGACCATGGCACTGACTCGAAGCACATAAAGTAGGTGTGACTTTTAATGCACATGG AAGAATCAAAACAAATGGATCTCTACTGAACACTTAACAGAGCATGAATGGGAACAG ATTATGATATTTGTGGAGTGATCCCAGAGGCAAAAATGGCTGTTTTCCAAATACGT GCCGAGGAGGGGGCTGCTATTTTGACCAGATGTTACTTCCAAGATTCTTAATAATA CCAGTTGAAGATGCTCATCAGGTCTCATGAATGTAAGCCCGAAGGGTATGAAATCTGT CATGATGGGAAGGTGGTACTATATTTCTGCTTCTAATTATTATGAGAAGGCAGCA ATCGAGGAGCTTACATCAAACATATGTTCTGGTACAACCTCCTCGATTTTTCCAGTACCA AGTAACATAAGCAACGTGCTTTTCAGCCTCTTCGCCAAAGAGTGGATACTATGGAAC AGCGCCATCAAGATATTAAGAGAGAGAGTGATTTACAGAAAAAGTGACCTTACTCGTG CTTTCCAACCTTCAAGACCACAGAAAAATCAGGAAAACTTTCTGTGAGCCAGTGGGCTTT TTGCATGGAGAACATTTTGGGGCTGAACTTACCATGGAGATCCCTCAGTTCGAATCTG GTAAACATAGACCAAAATGGAACGTTGAATACATGTCCAGCTTCCAGAATATCCGCA TTGAAAAACCTGTACAAGAGGCTCATTCTACTCTAGTTGAACTCTGTACAGATACAG		

	ATCTGACCTGGAAATCATATTTAATGCCATTGACACTGATCACTCAGGCCTGATCTCC GTGGAAGAATTTCTGTCATGTGGAACTTTTAGTTCTCACTACAATGTTTACATTG ATGATTCCCAAGTCAATAAGCTTGCCAACATAATGGACTTGAACAAAGATGGAAGCAT TGACTTTAATGAGTTTTTAAAGGCTTTCTATGTAGTGCATAGATGAAGACTTGATG AAACCTGATGTCACCAACCTTGGCTAAACACAAATGAGAGCTTCCCTCAGGCTCCCTG AAACAGCTAGGCCCAATCACAAGTACAGTCCTTTCCAACACCCCTGAAATTCATAGT CAGTAGCAG		
	ORF Start: ATG at 100		ORF Stop: TAA at 1939
	SEQ ID NO: 326	613 aa	MW at 71315.2kD
NOV35b, CG157629-01 Protein Sequence	MGCSSSSTKTRSDTSLRAALI IQNWYRGYKARLKARQHYALTI FQSI EYADEQGQMQ LSTFFSFMLENYTHIHKEELELRNQSLSEQDMRDRWDYVDSIDVPDSYNGPRLQFPL TCTDIDLLEAFKEQQILHAHYVLEVLFPETKKVLKQMPNFTHIQTSPEKVTTCGDLH GKLDDLFLIFYKNGLPSEPNPYVFNDFVDRGKNSIEILMILCVSFLVYPNDLHLNRG NHEDFMMNLRYGFTKEILHKYKLHGKRILQILEEFYAWLPETNDRHGTDSKHNVGV TFNAHGRIKTNGPSPEHLTEHEWEQIIDILWSDPRGKNGCFPNTCRGGGCFYFGPDVTS KILNKYQLKMLIRSHECKPEGYEICHGKVVTFISASNYEEGSRNGAYIKLCSGTTT RFFQYQVTKATCFQPLRQRVDTMENSAILLRERVISRKSDLTRAFQLQDHRKSGKLS VSQWAFCMENILGLNLPWRSLSNLVNIDQNGNVEYMSSFQNI RIEKPVQEAHSTLVE TLRYRSDLEIIFNAIDTDHSGLISVEBFRAMWKLFSHYNVHIDDSQVNKLANIMDL NKDGSIDFNEFLKAFYVVHRYEDLMKPDVTNLG		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 35B.

Table 35B. Comparison of NOV35a against NOV35b.		
Protein Sequence	NOV35a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV35b	1..613	613/613 (100%)
	1..613	613/613 (100%)

Further analysis of the NOV35a protein yielded the following properties shown in Table 35C.

Table 35C. Protein Sequence Properties NOV35a	
PSort analysis:	0.8171 probability located in mitochondrial matrix space; 0.4962 probability located in mitochondrial inner membrane; 0.4962 probability located in mitochondrial intermembrane space; 0.4962 probability located in mitochondrial outer membrane
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV35a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 35D.

Table 35D. Geneseq Results for NOV35a				
Geneseq	Protein/Organism/Length (Patent)	NOV35a	Identities/	Exact

Identifier	#, Date]	Residues/ Match Residues	Similarities for the Matched Region	Value
AAB47250	Human PP7 - Homo sapiens, 653 aa. [WO200130830-A2, 03-MAY-2001]	1..613 1..653	612/653 (93%) 612/653 (93%)	0.0
ABB71489	Drosophila melanogaster polypeptide SEQ ID NO 41259 - Drosophila melanogaster, 637 aa. [WO200171042-A2, 27-SEP-2001]	44..602 9..580	231/578 (39%) 341/578 (58%)	e-117
AAE09722	Novel cell cycle protein, protein phosphatase type 5 (PP5) - Unidentified, 499 aa. [WO200164913-A2, 07-SEP-2001]	86..422 156..487	126/343 (36%) 194/343 (55%)	3e-57
AAE09733	Protein phosphatase type 5 (PP5) variant, N303A - Unidentified, 499 aa. [WO200164913-A2, 07-SEP-2001]	86..422 156..487	125/343 (36%) 193/343 (55%)	2e-56
ABG09989	Novel human diagnostic protein #9980 - Homo sapiens, 500 aa. [WO200175067-A2, 11-OCT-2001]	86..422 160..491	125/343 (36%) 193/343 (55%)	3e-56

In a BLAST search of public sequence databases, the NOV35a protein was found to have homology to the proteins shown in the BLASTP data in Table 35E.

Table 35E. Public BLASTP Results for NOV35a				
Protein Accession Number	Protein/Organism/Length	NOV35a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
O14829	Serine/threonine protein phosphatase with EF-hands-1 (EC 3.1.3.16) (PPEF-1) (Protein phosphatase with EF calcium-binding domain) (PPEF) (Serine/threonine protein phosphatase 7) (PP7) - Homo sapiens (Human), 653 aa.	1..613 1..653	612/653 (93%) 612/653 (93%)	0.0
O01921	Hypothetical 80.3 kDa protein (Protein phosphatase with EF-hands) - Caenorhabditis elegans, 707 aa.	6..600 67..703	258/637 (40%) 375/637 (58%)	e-131
T34072	hypothetical protein F23H11.8 - Caenorhabditis elegans, 722 aa.	15..600 90..718	252/629 (40%) 368/629 (58%)	e-130
P40421	Serine/threonine nrotein phosnhatase	14..602	241/608 (39%)	e-123

	rdgC (EC 3.1.3.16) (Retinal degeneration C protein) - <i>Drosophila melanogaster</i> (Fruit fly), 661 aa.	3..604	360/608 (58%)	
AAM22065	<i>C. elegans</i> PEF-1 protein (corresponding sequence F23H11.8b) - <i>Caenorhabditis elegans</i> , 572 aa.	100..600 49..568	224/520 (43%) 319/520 (61%)	e-121

PFam analysis predicts that the NOV35a protein contains the domains shown in the Table 35F.

Table 35F. Domain Analysis of NOV35a			
Pfam Domain	NOV35a Match Region	Identities/ Similarities for the Matched Region	Expect Value
IQ	17..37	9/21 (43%) 17/21 (81%)	0.0022
STphosphatase	121..272	53/159 (33%) 115/159 (72%)	7.9e-46
STphosphatase	315..416	37/104 (36%) 83/104 (80%)	1.5e-34
efhand	530..558	12/29 (41%) 25/29 (86%)	3.4e-06
efhand	570..598	8/29 (28%) 24/29 (83%)	0.0011

### Example 36.

5 The NOV36 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 36A.

Table 36A. NOV36 Sequence Analysis			
	SEQ ID NO: 327	4037 bp	
NOV36a, CG157704-01. DNA Sequence	TTCACCAAAATGGCATCCTGGTTATATGAATGTCTTTGTGAAGCTGAAGTGCACAGT ATTATTCTCATTTCAGTCCCTTGGCCTTCAGAAAATAGATGAATTAGCCAAGATTAC AATGAAGGACTACTCCAAATTAGGAGTCCATGACATGAACGACCGCAAACGTCCTTC CACTTATCAAAATTATTAAGATTATGCAAGAAGAAGATAAAGCAGTCAGTATCCCAG AGCGTCATCTTCAGACAAGCAGCCTGCGCATCAAATCTCAGGAATTAAGATCTGGCCC TCGCAGACAGCTGAATTTTGATTCTCCTGCTGACAATAAAGACAGAAATGCCAGCAAT GATGGGTTTGAAATGTGCAGTTTATCAGATTTCTCTGCAAATGAACAGAAGTCCACTT ACCTAAAAGTGCTAGAACACATGCTACCAGATGATCCCAGTACCATACAAAAACAGG AATTCTGAATGCCACAGCTGGTGATTCTTATGTGCAAAACAGAAATCAGCACTTCACTC TTTTCACCAAAATTACCTTTCTGCAATACTGGGGGATTGTGATATTCCATTATTCAAA GAATCTCTCATGTTTCAGGTATAACTATGGAATCCCTCATTCTTGTATCAGACAGAA CACTTCAGAGAAAACAGAAATCCTTGGACTGAGATGGAGAAAATCAGAGTTTGTGTTCGA AAACGCCCCCTGGGCATGAGGGAGGTACGTCGTGGAGAAATTAATATTATTACTGTAG AAGACAAAGAACTCTACTTGTGCATGAGAAGAAAGAGCAGTTGACCTCACTCAATA TATTCTGCAGCATGTTTTTATTGATGAAGTCTTTGGTGAGGCGTGACCAATCAG		

	<p>GATGTATACATGAAGACTACTCACCACCTTATTTCAGCATATTTTCAATGGAGGCAATG  CCACTTGCTTTGCTTATGGACAGACAGGTGCTGGAAAGACCTACACCATGATAGGAAC  TCATGAGAACCAGGATTGTATGCTCTAGCTGCCAAAGATATCTTCAGGCAACTAGAA  GTGTCCAGCCAAGAAAGCACCTCTTTGTGTGGATCAGCTTCTAGAAATTTACTGTG  GACAGCTTTATGACCTCCTAAATAGAAAGAAAAAGGCTCTTTGCAAGAGAAGATAGCAA  GCACATGGTGCAGATAGTGGGACTGCAAGAGCTTCAGGTGGACAGTGTGGAGCTCCTC  TTACAGGTGATCTTAAAGGGCAGCAAGGAGCGCAGCACTGGGGCCACTGGAGTTAATG  CAGACTCCTCCCGCTCCCATGCCGTCATCCAAATTCAGATCAAAGATTACGCCAAGAG  GACATTTGGCAGGATCTCTTTTATTGACTTGGCTGGCAGTGAAAGAGCAGCAGATGCA  AGGGACTCAGATAGACAGACAAAGATGGAAGGTGCAGAAATAAATCAGAGTCTACTGG  CTCTGAAGGAATGTATCCGAGCACTGGATCAGGAACACACCCATACTCCCTTCAGGCA  AAGCAAATAACTCAGGTCTGAAGGACTCTTTCATCGGCAATGCCAAAACCTGCATG  ATCGCCAACATCTCACCAGCCACGTGGCCACTGAACACACTCTCAACACCTTGCGCT  ATGCTGACCGGGTCAAAGAACTAAAGAAAGGCATTAAGTGTGCACTTCAGTTACCAG  TCGAAATCGGACATCTGGAACCTCTCTCCAAAACGAATTCAGAGCTCCCTGGGGCT  TTGTGAGAGGACAAATGTCTCTCCAAAAAAGTCAAGCTGGGATTTCAGCAGTCACTCA  CAGTGGCAGCCCTGGTTCCACGAGAGGGAAGGTCCATCTCTGACCAGCCACCCACC  CAACATTCCTTTTACTTCTGCACCTAAGGTCTCTGGTAAAAGGGTGGCTCCAGAGGG  AGTCTTTCACAAGAGTGGGTCAATTCATGCTAGCCCTGTGAAAGGAACTGTGCGCTCTG  GACATGTGGCCAAAAAAGCCAGAAGAGTCAGCACCATTGTGCTCTGAGAAAAATCG  AATGGGCAACAAACTCTCCTTGGGTGGGAAAGCAGGGCCCTCAGGCCCAGGAGAAGGC  CTAGTGCCTGGTAAGCTGTCCACCAAGTGCAAGAAAGTGACAGCAGTGCAGCCAGTAC  AGAAGCAGCTTGTGTCTCGAGTTGAGCTCTCCTTTGGCAACGCCACCACAGGGCTGA  GTACAGTCAAGACAGCCAGAGGGGCGCCTGCTAGGCCTGCCTCTGAAGCTTGGACA  AACATCCCGCCACATCAGAAGGAGAGGGAGGAACATCTGCGTTTCTATCACCAGCAGT  TCCAACAGCCACCTCTCCTCCACAGAAGTTAAAAATACCAACCACTGAAAAGGTCTTT  ACGCCAGTACAGGCCCCAGAGGGTCAAGTCAAGAAAGAGACTCCGCCTCTGTTCAC  TCTTACTCTGAAAACCATGATGGAGCCCAAGTAGAGGAACCTGATGACAGTGAATTC  GTGAAGATTCTTTTTCACACATCTCTAGTCAGAGGGCCACAAAGCAAAGGAACCCCT  GGAGAATAGCGAAGACTCATTCTTCTGACACAGCTGGGGACAGGGTCTCTGAGAAG  CAGGTGGCAGAAAGACAGCAGAGTCTGTTTCTAGCCCCAGGACAGGTGACAAGAAAG  ATCTAACTAAAAGCTGGGTGGAATCCAGGGACCCCATAAACCACAGAAGAGCAGCACT  CGATCACAGCTGCAGCCCAAGTAAGGGGCCCCGTGGACTGGAGCAGAGAGAACTTACT  TCCTCAGGGCTTCTCCAGAGACAGCTGGCAGAGAAGCCATCTGTTTACAGGTAG  ATTTTCATATATAGACAGGAAAGAGGTGGAGGCTCTTCTTTGATCTCAGAAAGGATGC  CTCCCAAAGTGAGGTTTCTGGGGAGAATGAGGGCAACTTGCCATCCCCAGAGGAAGAT  GGTTTCACTATCTCATTGTCCACGTTGCAGTTCTTGATCCCCAGACCAAAGAGACA  CAGTCACCACACCTCTGAGAGAAGTCAAGTGCAGACGGCCCAATCCAGGTGACCAGCAC  TGTGAAAACGGTCTGCTGTCTCCAGGAGAGGATCTTAGGGGGCAGTTAGGCACGCAT  GCTGAATATGCTTCTGGACTCATGTCTCCCTCACCATGTCCCTCTGGAGAACCCAG  ACAACGAAGGGTCTCTCCCTCGGAGCAGCTGGTCCAGGATGGGGCTACGCACAGTCT  AGTGGCAGAGAGCACAGGGGGCCAGTTGTGAGCCACACAGTGCATCTGGTGTATCAA  GAGGCAGCCTTGCCAGTGTCTTCAGCAACTAGGCACCTGTGGCTGTCTCATCTCCCC  CTGATAATAAGCCTGGTGGTGATCTTCCAGCTCTGTCCCCATCACCATCCGTCAGCA  CCCAGCTGACAAGCTGCCAGCAGGGAGGCAGACCTAGGAGAGGCCTGCCAGAGCAGA  GAGACTGTACTTTTCTCCACGAACACATGGGTAGTGAGCAGTATGATGCTGATGCAG  AGGAGACGGGGCTGGATGGCTCTTGGGGTTTCCAGGAAAGCCCTTACCACCATACA  TATGGGGGTACCCCAATTCTGGACCTACACTACCCCAAGAACAGGAAGTAGTGATGTG  GCTGACCAGCTCTGGGCCAGGAGAGAAAAACATCTACAAGGCTTGGTTGGCAGGAGT  TTGGTTTGTCCACAGACCCCATCAAGTTGCCCTGCAACAGTGAAATGTACATGGCT  CAAACCCAGGCCGATCTCAAGGCAGGTGGTTCATCCGAGCACACCAGGAACAGCTGGAT  GAAATGGCTGAGCTCGGCTTCAAGGAGGAGACGCTGATGAGCCAGCTGGCTTCTAATG  ATTTTGAAGATTTTGTGACCCAGCTGGATGAAATCATGGTTCTGAAATCCAAGTGTAT  CCAGAGTCTGAGGAGCCAGCTGCAGCTCTATCTCACCTGCCACGGGCCACCGCAGCC  CCTGAGGGAACAGTGCCGCTTAGAGCCAGACCT</p>
	<p>ORF Start: ATG at 10</p>
	<p>ORF Stop: TAG at 4024</p>
	<p>SEQ ID NO: 328</p>
	<p>1338 aa</p>
	<p>MW at 148781.1kD</p>
NOV36a, CG157704-01	<p>MASWLYECLCEAELAQYSHFTALGLQKIDELAKITMKDYSKLGVHDMNDRKRLFQLI  KIIKIMQEEDKAVSIPERHLQTSRLRIKSQELRSGPRRLNFDSPADNKDRNASNDGF</p>

Protein Sequence	EMCSLSDFSANEQKSTYLKVLEHMLPDDSQYHTKTGILNATAGDSYVQTEISTSLFSP NYLSAILGDCDIP I IQRISHVSGYNYGI PHSCIRQNTSEKQNPWTEMEKIRVCVRKRP LGMREVRERGEINIITVEDKETLLVHEKKEAVDLTQYILQHVFFYFDEVFGEACTNQDVY MKTTHPLIQHIFNGGNATCFAYGQTGAGKTYTMIGTHENPGLYALAAKDI FRQLEVSQ PRKHLFVWISFYEIYCGQLYDLLNRRKRLFAREDSKHMVQIVGLQELQVDSVELLLQV ILKGSKERSTGATGVNADSSRSHAVIQIQIKDSAKRTFGRISFIDLAGSERAADARDS DRQTKMEGAEGINQSLALKECIRALDQEHHTHPFRQSKLTQVLKDSFIGNAKTCMIAN ISPSHVATEHTLNTLRYADRVKELKKGIKCCTSVTSRNRTSGNSSPKRIQSSPGALSE DKCSPKKVKLGFQQSLTVAAPGSTRGKVHPLTSHPPNIPFTSAPKVS GKRGGSRGSPS QEWVIHASPVKGTVRSGHVAKKKPEESAPLCSEKNRMGNKTVLGWESRASGPGEGLVR GKLSTKCKKVQTVQPVQKQLVSRVELSFGNAHRAEYSQDSQRGTFARPASEAWTNIP PHQKEREHLRFYHQFQQPPLLQQLKYQPLKRLRQYRPPEGLTNETPPLFHSYS ENHDGAQVEELDDSDFSSEDSFSHISSQRATKQRNTLENSEDSFFLHQTWQGQPEKQVA ERQQSLFSSPRTGDKDLTKSWVDSRDPINHRRALDHSCSPSKGPDVDSRENSTSSG PSPRDSLAEKPYCSQVDFIYRQERGGSSFDLRKQASQSEVSGENEGNLPSPEEDGFT ISLSHVAVPGSPDQDRTVTTPLEVSADGPIQVTSTVKNGHAVPGEDPRGQLGTHAEY ASGLMSPLTMSLLENPDNEGSPSEQLVQDGATHSLVAESTGGPVVSHTPVPSGDQEAAL LPVSSATRHLWLSSSPDNKPGGDLPALSPSPIRQHPADKLPREADLGEACQSRQTV LFSHEHMGSEQYDADAETGLDGSWGFPGKPF'TTIHMGVPHSGPTLTPTRTGSSDVADQ LWAQERKHPTRLGWQEFGLSTDPIKLPCNSENVTLKPRPISRQVVI RAHQEQLDEMA ELGFKEETLMSQLASNDFEDFVTQLDEIMVLKSKCIQSLRSQQLYLTCHGPTAAPEG TVPS
------------------	---

Further analysis of the NOV36a protein yielded the following properties shown in Table 36B.

Table 36B. Protein Sequence Properties NOV36a	
PSort analysis:	0.8200 probability located in nucleus; 0.3000 probability located in microbody (peroxisome); 0.1000 probability located in mitochondrial matrix space; 0.1000 probability located in lysosome (lumen)
SignalP analysis:	No Known Signal Sequence Predicted

5 A search of the NOV36a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 36C.

Table 36C. Geneseq Results for NOV36a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV36a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAU77182	Human kinesin motor protein KinI-3 - Homo sapiens, 1368 aa. [WO200226929-A2, 04-APR-2002]	1..1338 1..1368	1337/1368 (97%) 1338/1368 (97%)	0.0
AAU77184	Human KinI-3 DNA fragment with flanking vector sequences #2 - Homo sapiens. 381 aa.	195..566 2..373	371/372 (99%) 372/372 (99%)	0.0

	[WO200226929-A2, 04-APR-2002]			
AAU77183	Human KinI-3 DNA fragment with flanking vector sequences #1 - Homo sapiens, 373 aa. [WO200226929-A2, 04-APR-2002]	183..546 2..365	363/364 (99%) 364/364 (99%)	0.0
AAU77186	Human KinI-3 DNA fragment with flanking vector sequences #4 - Homo sapiens, 363 aa. [WO200226929-A2, 04-APR-2002]	213..566 2..355	353/354 (99%) 354/354 (99%)	0.0
AAU77185	Human KinI-3 DNA fragment with flanking vector sequences #3 - Homo sapiens, 343 aa. [WO200226929-A2, 04-APR-2002]	213..546 2..335	333/334 (99%) 334/334 (99%)	0.0

In a BLAST search of public sequence databases, the NOV36a protein was found to have homology to the proteins shown in the BLASTP data in Table 36D.

Table 36D. Public BLASTP Results for NOV36a				
Protein Accession Number	Protein/Organism/Length	NOV36a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q9GYC7	Probable mitotic centromere associated kinesin - Leishmania major, 728 aa.	1..548 1..519	222/551 (40%) 317/551 (57%)	e-101
Q9NV43	OVARC1000605 protein - Homo sapiens (Human), 172 aa.	37..208 1..172	172/172 (100%) 172/172 (100%)	5e-95
Q94GW1	Kinesin-like protein - Oryza sativa (Rice), 800 aa.	208..574 188..539	192/368 (52%) 251/368 (68%)	3e-94
P28740	Kinesin-like protein KIF2 - Mus musculus (Mouse), 716 aa.	223..617 195..582	196/407 (48%) 259/407 (63%)	2e-93
Q9VZ28	CG1453 protein - Drosophila melanogaster (Fruit fly), 803 aa.	223..546 276..608	182/333 (54%) 236/333 (70%)	5e-93

PFam analysis predicts that the NOV36a protein contains the domains shown in the Table 36E.

Table 36E. Domain Analysis of NOV36a
--------------------------------------

Pfam Domain	NOV36a Match Region	Identities/ Similarities for the Matched Region	Expect Value
SAM	2..62	19/68 (28%) 41/68 (60%)	0.42
kinesin	229..547	129/388 (33%) 236/388 (61%)	3.7e-89

**Example 37.**

The NOV37 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 37A.

Table 37A. NOV37 Sequence Analysis			
	SEQ ID NO: 329	2770 bp	
NOV37a, CG158218-01 DNA Sequence	TTATGGGACCATGATGTTGAGAGTTAGTGTGAAGTGGACCATGAAAAAGCCAGCCAA AGTAGCATCTTCATCCGTTTCCAGGCCATGCCCTTTCATTATAACCAGAAGGCCCCAT GTTCTGAGTGCCATGATCTGATGTGTAGGAATGTCAATCCACCCCGCAGATCATTGC AACTTTAGTGGACATACCTATACATGCCAAAAGCATCCTGCCTCCAGGGTCTGCACCT CTCTCTGCCCAACGGCTTCTCTGAATGTCAGGGCACACAGGATTTATTCATAGATG AAGATGAAAAATTAATACCTAGCTTGGAAATCATCTTACCACGTGATTGGCAGATGG GTTTGTGAATAATAAGCGAGAAAGCTACAAATTTAAATTTCAAAGAATTTTGTATCAG GATGCAAACCAAGAGACCGTTTTTGAACACATTGCCAAACCAAGTTGCTGGGAGGTATC TCACCCCTGGTGGTAAGGATGTCTGGCAGGTTACAATGGTACCATCTTTGCATATGG GCAAACAGGCAGCGGAAGACATTCACTATCACAGGGGGTGACAGAGCGTTACAGTGAC AGAGGCATTATCCCAAGGACACTGTCATACATTTTGAACAGTTACAAAAGGACAGCA GCAAAATATATACACACACATTTCCCTATTTGGAATCTACAATGAATGTGGTTATGA TCTTTTGGATCCAAGACATGAAGCCTCCAGTTTGAAGATTTGCCGAAAGTGACAATA CTGGAGGATCCTGTACAGAACATTCACCTGAAAACTTGACTCTCCATCAGGCAACCA CAGAGGAAGAAGCTCTGAATTTGCTTTTTTTAGGAGACACCAACCGAATGATTGCAGA GACTCCTATGAACCAAGCTTCAACCCGTTCCCACTGCATTTTACCATTTCATTTGTCA AGCAAAGGAACAGGATCTGCAACTGTACGACATGCCAAACTCCATCTGGTTGACCTGG CTGGTTCAGAGCGAGTTGCAAAGACTGGAGTAGGGGGCCATCTTCTAACAGAGGCCAA GTATATCAACTTGTCACTACATTACTTAGAACAGGTTATCATTGCCCTTTCAGAAAAG CACCGTTCGCACATTCCTTATAGAACTCCATGATGACCAAGTGTCTTAAGAGACAGTT TGGGAGGGAAGTGCATGACAACATGATTGCAACACTCTCCTTGGAGAAAAGGAATCT TGATGAGTCTATATCAACCTGCAGATTTGCACAGCGAGTGGCACTATAAGAATGAA GCTGTCTTAAATGAAGAAATTAACCCAGATTAGTGATTAAACGCCTACAAAAGGAAA TCCAGGAAGTGAAGGATGAAGTGGCCATGGTCACTGGGGAGCAGAGGACAGAGGCACT CACAGAAGCAGAGCTCCTTCAGCTGGAAAACTAATAACATCCTTTTTGGAAGACCAG GATTCAGACAGTAGATTAGAGGTTGGCGCGGATATGCGTAAAGTTTCACTGTTTTTC ATCATTTAAAGAACTATTGAATGACAAGAAGATCCTTGAAAAACAATACAGTCTCCTC TGAAAGCAAAGACCAAGATTGTCAAGAACCATTAAAGAAGAAGAATATAGAAAGCTA CGAGATATTCTGAAACAGAGAGATAACGAAATCAATATCCTGGTCAACATGTTAAAAA AAGAAAAGAAGAAAGCTCAGGAGGCTCTCCACTTGGCTGGCATGGATAGACGTGAATT CAGACAGTCCCAGAGCCACCTTCCCGCTAGGAAACCCAGAAGAAGGTCAAAGAATG CGACTATCCTCAGCTCCCTCACAGGCCAGGACTTCAGCATTTTGGGGAAAAGATCCA GTTTGCTCCACAAGAAAATAGGAATGAGAGAGGAAATGTATTAGGATGCCAGGAGGC TTTTGAAATCTTCAAGAGGGACACGCTGACAGCGTTACCATCGATGACAAACAAACAG ATTCTGAAACAGAGATTTTCTGAAGCCAAGGCCCTGGGAGAAAGTATAAATGAAGCAA GAAGTAAAATTTGGTCACTGAAGGAAGAAATCACCCAGCGGCATATACAGCAAGTAGC CCTAGGAATCTCGAAAAACATGGCCGTGCCTCTGATGCCAGACCAGCAGGAGGAGAAG CTGCGATCACAACTGGAGGAAGAAAAGAGAAGGTATAAAACAATGTTCACTCGCCTGA AAGCCTGAAGGTGGAGATCGAGCACTTGACGCTGCTCATGGACAAAGCCAAGGTGAA		

	GCTACAGAAAGAGTTTGAAGTCTGGTGGGCAGAGGAGGCCACCAACCTGCAGGTAAAT TCTCCAGCAGTGAATTCACCTCGATCACACGAAGCCATTTCTCCAGACATCTGACTCCC AGCATGAATGGTCCCACTCCTCTTAACAAAAGTTCTGGAGGCTGGGAAGTCCAAGA TCAAGGCACTGGCAGATTTCGATGTCTGTGATGTGAATGCCAGGAAAATCCTGCCCTCG CCTTGCCCCAGTCCACACAGCCAGAAACAGAGCAGCACCAGCACCCTGGAAGACA GCATCCCCAAGAGGCCAGTGTCTGTCATCCCTCTCACCAGGAGACAGCCAGCAGGACTC GGACATCATCGCCTTCATCAAGGCCAGACAGAGCATTCTGCAGAAGCAATATCTTCAG CTCCTTTGTTCTCTGTTCCCAAAGTCAGCTGTCTCTCTGCTCAGGCTTCTACAAACA GGAAGGGGCTGAGTGATGTTTTGGTAACTCGTTGAACCCCTGGC		
	ORF Start: ATG at 11		ORF Stop: TGA at 2759
	SEQ ID NO: 330	916 aa	MW at 103840.1kD
NOV37a, CG158218-01 Protein Sequence	MMLRVSVKWTIEKASQSSIFIRFQAMPFHYNQKAPCSECHDLMCRNVNSTPQIIATLV DIPIHAKSILPPGSAPLSAQRLSLNVRHRIYSIDEDEKLIPSLIILPRDLADGFVN NKRESYKFKFQRIFDQDANQETVFENIAKPVAGRYLTPGGKDVLAGYNGTIFAYGQTG SGKTFTITGGAERYSDRGIIIPRTLSYIFEQLQKDSKIYTHISYLEIYNECGYDLLD PRHEASSLEDLPKVITILEDPDQNIHLKNTLHQATTEEEALNLLFLGDTNRMIAETPM NQASTRSHCIFTHLSSKEPGSATVRHAKLHLVDLAGSERVAKTGVGGHLLTEAKYIN LSLHYLEQVIIALSEKHRSHI PYRNSMNTSVLRDSLGGNCMTMIATLSLEKRNLDSE ISTCRFAQRVALIKNEAVLNEEINPRLVIKRLQKEIQELKDELAMVTGEORTEALTEA ELLQLEKLITSFLEDQSDSRLEVADMRKVHCFHHLKKLLNDKKILENNTVSSSESK DQDCQEPLKEEEYRKLRLDILKQRDNEINILVNMMLKKEKKKAQEALHLAGMDRREFRQS QSPFRLGNPEEGQRMRLSSAPSQAQDFSILGKRSSLHKKIGMREEMSLGCQEAFEI FKRDHADSVTIDDNKQILKQRFSEAKALGESINEARSKI GHLKEEITQRHIQQVALGI SENMAVPLMPDQEEKLRSQLEEEKRRYKTMFTRLKALKVEIHLQLLMDKAKVKLQK EFEVWAAEEATNLQVNSPAVNSLDHTKPFQTSDSQHEWSQLLSNKS SGGWEVQDQGT GRFDVCDVNARKILPSPCPSPHSQQSSTSTPLEDSIPKRPVSSIPLTGDSQTDSDIT AFIKARQSILQKQYLQLLCSLFPKSAVSSAQASTNRKGLSDVLVTR		

Further analysis of the NOV37a protein yielded the following properties shown in Table 37B.

Table 37B. Protein Sequence Properties NOV37a	
PSort analysis:	0.6863 probability located in mitochondrial matrix space; 0.3737 probability located in mitochondrial inner membrane; 0.3737 probability located in mitochondrial intermembrane space; 0.3737 probability located in mitochondrial outer membrane
SignalP analysis:	No Known Signal Sequence Predicted

- 5 A search of the NOV37a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 37C.

Table 37C. Geneseq Results for NOV37a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV37a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAU75177	Human kinesin protein 9 - Homo saniens. 725 aa. [CN1319665-A.	86..762 20..643	220/685 (32%) 363/685 (52%)	3e-91

	31-OCT-2001]			
AAE14609	Human microtubule motor protein HsKif6 motor domain - Homo sapiens, 205 aa. [US6346410-B1, 12-FEB-2002]	159..322 28..191	164/164 (100%) 164/164 (100%)	3e-91
AAU75800	Human ortholog of mouse kinesin Kif9, HsKif9 - Homo sapiens, 790 aa. [US6331430-B1, 18-DEC-2001]	86..762 20..708	217/739 (29%) 362/739 (48%)	8e-81
ABB80741	Human kinesin motor protein, HsKif9 sequence - Homo sapiens, 790 aa. [US6355447-B1, 12-MAR-2002]	86..762 20..708	217/739 (29%) 362/739 (48%)	8e-81
AAB94768	Human protein sequence SEQ ID NO:15849 - Homo sapiens, 664 aa. [EP1074617-A2, 07-FEB-2001]	86..510 20..433	162/432 (37%) 258/432 (59%)	1e-77

In a BLAST search of public sequence databases, the NOV37a protein was found to have homology to the proteins shown in the BLASTP data in Table 37D.

Table 37D. Public BLASTP Results for NOV37a				
Protein Accession Number	Protein/Organism/Length	NOV37a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
O54720	Kinesin-related protein 3A - Rattus norvegicus (Rat), 486 aa (fragment).	81..560 15..486	416/480 (86%) 442/480 (91%)	0.0
Q8R471	Kinesin-related protein 3B - Rattus norvegicus (Rat), 452 aa (fragment).	81..507 14..432	376/427 (88%) 396/427 (92%)	0.0
Q8WTV4	Hypothetical 30.1 kDa protein - Homo sapiens (Human), 265 aa.	624..885 1..262	261/262 (99%) 261/262 (99%)	e-147
Q9UJR0	DJ1043E3.1 (Novel protein) - Homo sapiens (Human), 189 aa (fragment).	434..622 1..189	189/189 (100%) 189/189 (100%)	e-102
O35067	Motor domain of KIF6 - Mus musculus (Mouse), 165 aa (fragment).	167..329 1..165	155/165 (93%) 158/165 (94%)	2e-84

PFam analysis predicts that the NOV37a protein contains the domains shown in the Table 37E.

Table 37E. Domain Analysis of NOV37a			
Pfam Domain	NOV37a Match Region	Identities/ Similarities for the Matched Region	Expect Value
kinesin	124..449	153/375 (41%) 255/375 (68%)	6.5e-119

**Example 38.**

The NOV38 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 38A.

Table 38A. NOV38 Sequence Analysis			
	SEQ ID NO: 331	1184 bp	
NOV38a, CG158513-01 DNA Sequence	AGTTCCTCCTAACTCCTGCCAGAAACAGCTCTCCTCAACATGAGAGCTGCACCCCTCC TCCTGGCCAGGGCAGCAAGCCTTAGCCTTGGCTTCTTGTTCTGCTTTTTTCTGGCT AGACCGAAGTGTA TAGCCAAGGAGTTGAAGTTGTGACTTTGGTGTTCGGCATGGA GACCGAAGTCCCATTGACACCTTTCCCACTGACCCATAAAGGAATCCTCATGGCCAC AAGGATTTGGCCAACCTACCCAGCTGGGCATGGAGCAGCATTATGAACCTGGAGAGTA TATAAGAAAGAGATATAGAAAATCTTGAATGAGTCCTATAAACATGAACAGGTTTAT ATTCGAAGCACAGACGTTGACCGGACTTTGATGAGTGCTATGACAAACCTGGCAGCCC TGTTTCCCCCAGAAGGTGTGAGCATCTGGAATCCTATCCTACTCTGGCAGCCCATCCC GGTGACACAGTTCTCTTTCTGAAGATCAGGATTTTATAGCTACCTTGGGAAAACCTT TCAGGATTACATGGCCAGGACCTTTTGGAAATTTGGAGTAAAGTCTACGACCCCTTAT ATTGTGAGAGTGTTCACAATTTCACTTTACCCCTCCTGGGCCACTGAGGACACCATGAC TAAGTTGAGAGAATTGTGAGAATTGTCCCTCCTGTCCCTCTATGGAATTCACAAGCAG AAAGAGAAATCTAGGCTCCAAGGGGGTGTCTGGTCAATGAAATCCTCAATCATGTA AGAGAGCAACTCAGATACCAAGCTACAAAAAATCATCATGTATTCTGCGCATGACAC TACTGTGAGTGGCTACAGATGGCGCTAGATGTTTACAACGGACTCCTTCTCCCTAT GCTTCTTGCCACTTGACGGAATTGTACTTTGAGAAGGGGGAGTACTTTGTGGAGATGT ACTATCGGAATGAGACGACGACGACCGGTATCCCTCATGCTACCTGGCTGCAGCCC CAGCTGTCCTCTGAGAGAGTTTGTGAGCTGGTTGGCCCTGTGATCCCTCAAGACTGG TCCACGGAGTGATGACCACAAACAGCCATCAAGGTACTGAGGACAGTACAGATTAGT GTGCACAGAGATCTCTGTAGAAAGAGTAGCTGCCCTTTCTCAGGGCAGATGATGCTTT GAGAACATACTTTGGCCATTACCC		
	ORF Start: ATG at 40		ORF Stop: TAG at 1099
	SEQ ID NO: 332	353 aa	MW at 40442.9kD
NOV38a, CG158513-01 Protein Sequence	MRAAPLLLARAASLSLGLFLLFFWLDRSVLAKELKFVTLVFRHGDSPIDTFPTDPI KESWQFGQQLTQLGMEQHYELGEYIRKRYRKFLNESYKHEQVYIRSTDVRTLMSA MTNLAALFPPEGVSIWNPI LLWQPI PVHTVPLSEDQDFIATLGLKSLHGQDLFGIWS KVYDPLYCESVHNFTLPWATEDMTKLRELSEL SLLSLYGIHKQKEKSRLQGGVLVN EILNHMKRATQIPSYKKLIMYSAHDTTVSGLQMALDVYNGLLPPYASCHLTelyFEKG EYFVEMYRNETQHEPYPLMLPGCSPSPCLERFAELVGPVLPQDWSTECMTTNSHQGT EDSTD		
	SEQ ID NO: 333	1184 bp	
NOV38b, CG158513-02 DNA Sequence	AGTTCCTCCTAACTCCTGCCAGAAACAGCTCTCCTCAACATGAGAGCTGCACCCCTCC TCCTGGCCAGGGCAGCAAGCCTTAGCCTTGGCTTCTTGTTCTGCTTTTTTCTGGCT AGACCGAAGTGTA TAGCCAAGGAGTTGAAGTTGTGACTTTGGTGTTCGGCATGGA GACCGAAGTCCCATTGACACCTTTCCCACTGACCCATAAAGGAATCCTCATGGCCAC AAGGATTTGGCCAACCTACCCAGCTGGGCATGGAGCAGCATTATGAACCTGGAGAGTA TATAAGAAAGAGATATAGAAAATCTTGAATGAGTCCTATAAACATGAACAGGTTTAT ATTCGAAGCACAGACGTTGACCGGACTTTGATGAGTGCTATGACAAACCTGGCAGCCC		

	TGTTCCTCCCAAGGTGTCAGCATCTGGAATCCTATCCTACTCTGGCAGCCCATCCC GGTGCACACAGTTCTCTTTCTGAAGATCAGGATTTTATAGCTACCTTGGGAAACTT TCAGGATTACATGGCCAGGACCTTTTGGGAATTGGAGTAAAGTCTACGACCTTTAT ATTGTGAGAGTGTTCACAATTTCACTTTACCCTCCTGGGCCACTGAGGACACCATGAC TAAGTTGAGAGAATTGTCAGAATTGTCCCTCCTGTCCCTCTATGGAATTCACAAGCAG AAAGAGAAATCTAGGCTCCAAGGGGGTGTCTGCTCAATGAAATCCTCAATCACATGA AGAGAGCAACTCAGATACCAAGCTACAAAAAATTATCATGTATTCTGCGCATGACAC TACTGTGAGTGGTCTACAGATGGCGCTAGATGTTTACAACGGACTCCTTCTCCCTAT GCTTCTTGCCACTTGACGGAATTGTACTTTGAGAAGGGGGAGTACTTTGTGGAGATGT ACTACCGGAATGAGACGCAGCAGAGCCGTATCCCCTCATGCTACCTGGCTGCAGCCC CAGCTGTCTCTGGAGAGGTTTGCTGAGCTGGTTGGCCCTGTGATCCCTCAAGACTGG TCCACGGAGTGTATGACCACAAACAGCCATCAAGGTACTGAGGACAGTACAGATTAGT GTGCACAGAGATCTCTGTAGAAAGAGTAGCTGCCCTTTCTCAGGGCAGATGATGCTTT GAGAACATACTTTGGCCATTACCC		
	ORF Start: ATG at 40		ORF Stop: TAG at 1099
	SEQ ID NO: 334	353. aa	MW at 40442.9kD
NOV38b, CG158513-02 Protein Sequence	MRAAPLLARAASLSLGFLLFFWLDRLVLAKEKFVTLVFRHGDRSPIDTFPTDPI KESWPQGFQQLTQLGMEQHYELGEYIRKRYRKLNESYKHEQVYIRSTVDRTLMSA MTNLALFPPEGVSIWNPIILLWQPIPVHTVPLSEDDFIATLGKLSGLHGQDLFGIWS KVYDPLYCESVHNFTLPSWATEDTMTKLRELSELSLLSLYGIHKQEKSRLOGGVLVN EILNHMKRATQIPSYKKLIMYSAHDTTVSGLQMALDVYNGLLPPYASCHLTELYFEKG EYFVEMYRNETQHEFYPLMLPGCSPSCPLERFAELVGPVLPQDWSTECMTTNSHQGT EDSTD		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 38B.

Table 38B. Comparison of NOV38a against NOV38b.		
Protein Sequence	NOV38a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV38b	1..353	353/353 (100%)
	1..353	353/353 (100%)

5 Further analysis of the NOV38a protein yielded the following properties shown in Table 38C.

Table 38C. Protein Sequence Properties NOV38a	
PSort analysis:	0.4600 probability located in plasma membrane; 0.2083 probability located in microbody (peroxisome); 0.1000 probability located in endoplasmic reticulum (membrane); 0.1000 probability located in endoplasmic reticulum (lumen)
SignalP analysis:	Cleavage site between residues 33 and 34

A search of the NOV38a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 38D.

Table 38D. Geneseq Results for NOV38a
---------------------------------------

Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV38a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAB74820	Prostate tumour antigen amino acid sequence for PAP - Homo sapiens, 386 aa. [WO200125272-A2, 12-APR-2001]	1..353 1..386	353/386 (91%) 353/386 (91%)	0.0
AAG62145	Human prostatic acid phosphatase SEQ ID NO: 328 - Homo sapiens, 386 aa. [WO200125273-A2, 12-APR-2001]	1..353 1..386	353/386 (91%) 353/386 (91%)	0.0
AAU02172	Biomarker UC band 47 (PAP), used in diagnosis and prognosis of cancer - Homo sapiens, 386 aa. [US6218529-B1, 17-APR-2001]	1..353 1..386	353/386 (91%) 353/386 (91%)	0.0
AAU06277	Prostatic Acid Phosphatase (PAP) polypeptide - Homo sapiens, 386 aa. [WO200145728-A2, 28-JUN-2001]	1..353 1..386	353/386 (91%) 353/386 (91%)	0.0
AAY59293	Prostatic acid phosphatase marker UC Band #47 amino acid sequence - Homo sapiens, 386 aa. [WO9964631-A1, 16-DEC-1999]	1..353 1..386	353/386 (91%) 353/386 (91%)	0.0

In a BLAST search of public sequence databases, the NOV38a protein was found to have homology to the proteins shown in the BLASTP data in Table 38E.

Table 38E. Public BLASTP Results for NOV38a				
Protein Accession Number	Protein/Organism/Length	NOV38a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
P15309	Prostatic acid phosphatase precursor (EC 3.1.3.2) - Homo sapiens (Human), 386 aa.	1..353 1..386	353/386 (91%) 353/386 (91%)	0.0
Q96KY0	Acid phosphatase, prostate - Homo sapiens (Human), 386 aa.	1..353 1..386	352/386 (91%) 353/386 (91%)	0.0
Q96QK9	Acid phosphatase, prostate - Homo sapiens (Human), 386 aa.	1..353 1..386	350/386 (90%) 351/386 (90%)	0.0
Q96QM0	Acid phosphatase, prostate - Homo sapiens (Human), 418 aa.	1..346 1..379	345/379 (91%) 345/379 (91%)	0.0
Q9QXH7	Prostatic acid phosphatase - Mus musculus (Mouse), 381 aa.	1..347 1..379	281/380 (73%) 307/380 (79%)	e-162

Pfam analysis predicts that the NOV38a protein contains the domains shown in the Table 38F.

Table 38F. Domain Analysis of NOV38a			
Pfam Domain	NOV38a Match Region	Identities/ Similarities for the Matched Region	Expect Value
Acid_phosphat	33..340	128/436 (29%) 300/436 (69%)	2.7e-126

### Example 39.

5. The NOV39 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 39A.

Table 39A. NOV39 Sequence Analysis			
	SEQ ID NO: 335	1967 bp	
NOV39a, CG158583-01 DNA Sequence	GGAGCCATGGCCCTGAGCGAGCTGGCGCTGGTCCGCTGGCTGCAGGAGAGCCGCCGCT CGCGGAAGCTCATCCTGTTTCATCGTGTTCCTGGCGCTGCTGCTGGACAACATGCTGCT CACTGTCGTGGTAGAGAGAGGGTTTCTCCATGTTGGCCAGCCTGGTCTCGAACTCCTG ACCTCAGGTGATCCACCTGCCTCAGCTTCCCAAAGTCCGGAATTACAGTCCCCATCA TCCCAAGTTATCTGTACAGCATTAAGCATGAGAAGAATGCTACAGAAATCCAGACGGC CAGGCCAGTGCACACTGCCCTCCATCTCAGACAGCTTCCAGAGCATCTTCTCCTATTAT GATAACTCGACTATGGTCACCGGGAATGCTACCAGAGACCTGACACTTCATCAGACCG CCACACAGCACATGGTGACCAACGCGTCCGCTGTTCCCTCCGACTGTCCAGTGAAGA CAAAGACCTCCTGAATGAAACGCTGCAAGTTGGTCTGTGTTTGGCTCGAAAGCCACC GTCCAGCTCATACCAACCCCTTTCATAGGACTACTGACCAACAGAATTGGCTATCCAA TTCCCATATTTGCGGGATTCTGCATCATGTTTGTCTCAACAATTATGTTTGCCTTCTC CAGCAGCTATGCCTTCTGCTGATTGCCAGGTGCTGCGAGGCGCATCGCTCGTCTGCTG TCCTCTGTGGCTGGGATGGGCATGCTTGCCAGTGTCTACACAGATGATGAAGAGAGAG GCAACGTCATGGGAATCGCCTTGGGAGGCTGGCCATGGGGGTCTTAGTGGGCCCCC CTTCGGGAGTGTCTCTATGAGTTTGTGGGAAGACGGCTCCGTTCTGGTGTGCGCC GCCCTGGTACTCTTGGATGGAGCTATTGAGCTCTTGTGCTCCAGCCGTCGCCGGGTGC AGCCAGAGAGTCAGAAGGGGACACCCCTAACCACGCTGCTGAAGGACCCGTACATCCT CATTGCTGCAGGCTCCATCTGCTTTGCAAACATGGGCATCGCCATGCTGGAGCCAGCC CTGCCCATCTGGATGATGGAGACCATGTGTTCCCGAAAGTGGCAGCTGGGCGTTGCCCT TCTTGCCAGCTAGTATCTCTTATCTCATTTGGAACCAATATTTTGGGATACTTGCACA CAAAATGGGGAGGTGGCTTTGTGCTCTCTGGAATGATAATTGTTGGAGTCAGCATT TTATGTATTTCCATTTGCAAAAAACATTTATGGAATCATAGCTCCGAACCTTTGGAGTTG GTTTGTGCAATTGGAATGGTGGATTTCGTCAATGATGCCATCATGGGCTACCTCGTAGA CCTGCGGCACGTGTCCGTCTATGGGAGTGTGTACGCCATTGCGGATGTGGCATTTTGT ATGGGGTATGCTATAGGTCCTTCTGCTGGTGGTGTCTATGCAAAGGCAATTGGATTTC CATGGCTCATGACAATTATTTGGGATAATTGATATTTCTTTTGGCCCTCTCTGCTTTT TCTTCGAAGTCCACCTGCCAAAGAAGAAAAATGGCTATTCTCATGGATCACAACCTGC CCTATTAAAAACAAAATGTACACTCAGAATAATATCCAGTCATATCCGATAGGTGAAG ATGAAGAATCTGAAAGTGAAGTGAAGATCCTCAAAATCATCAAGTGTTTAATT GTATAAACAGTGTTCAGTGACACAACCTCATCCAGAAGTGTCTTAGTCATACCATC CATCCCTGGTGAAAGAGTAAACCAAAGGTATTATTTCCTTTCCATGGTTATGTCG ATTGCCAACAGCCTTATAAAGAAAAAGAGCTTTTCTAGGGGTTTGTATAAATAGTGT TGAAACTTTATTTATGTATTTAATTTTATTAATATCATACAATATATTTTGATGAA ATAGGTATTGTGTAATCTATAAATATTGAATCCAAACCAATATAATTTCC		
	ORF Start: ATG at 7		ORF Stop: TGA at 1645
	SEQ ID NO: 336	546 aa	MW at 58912.5kD

NOV39a, CG158583-01 Protein Sequence	MALSELALVRWLQESRRSRKLILFIVFLALLLDNMLLTVVVERGFLHVGQPGLELLTS GDPPASASQSPGITVPIIPSYLYSIKHEKNATEIQARPVHTASISDSFQSIFSYYDN STMVTGNATRDLLHQTATQHMVTNASAVPSDCPSEDKDLLNENVQVGLLFASKATVQ LITNPFIGLLTNRIGYPIPIFAGFCIMFVSTIMFAFSSSYAFLLIARSLQIGSSCSS VAGMGLASVYTDDEERGNVMGIALGGLAMGVLVGPPFGSVLYEFVGKTAPFLVLAAL VLLDGAIQFLVLPQSRVQPSQKGTPLTLLKDPYILIAAGSICFANMGIAMLEPALP IWMETMCSRKWQLGVAFLPASISYLIGTNIFGILAHKMGRWLCALLGMIIVGVSIILC IPFAKNIYGLIAPNFGVGFAGMVDSSMMPIMGYLVDLRHVSVYGSVYAIADVAFCMG YAIGPSAGGAIKAIGFPWLMTIIGIIDILFAPLCFFLRSPPAKEEKMAILMDHNCPI KTKMYTQNNIQSYPIGEDEESES		
	SEQ ID NO: 337	1952 bp	
NOV39b, CG158583-02 DNA Sequence	GCAGGCATCGCAAGCGACCCCGAGCGGAGCCCGAGCCATGGCCCTGAGCGAGCTGG CGCTGGTCCGCTGGCTGCAGGAGAGCCGCCGCTCGCGGAAGCTCATCTGTTCATCGT GTTCTTGGCGCTGCTGCTGGACAACATGCTGCTCACTGTCTGGGTTCAAGCGATCCT CCTTCTCAGCCTCCAAAGGAGCTGGGATTACAGTCCCATCATCCCAAGTTATCTGT ACAGCATTAAGCATGAGAAGAATGCTACAGAAATCCAGACGCCAGGCCAGTGACAC TGCCTCCATCTCAGACAGCTTCCAGAGCATCTTCTCCTATTATGATAACTCGACTATG GTCACCGGAATGCTACCAGAGACCTGACACTTCATCAGACGCCACACAGCATATGG TGACCAACCGCTCCGCTGTTCTTCCGACTGTCCAGTGAAGACAAAGACTCCTGAA TGAAAACGTGCAAGTTGGTCTGTTGTTTGCCTCGAAAGCCACCGTCCAGCTCATCACC AACCCTTTCATAGGACTACTGACCAACAGAATTGGCTATCCAATTCCCATATTGCGG GATTCTGCATACATGTTGTCTCAACAATTATGTTTGCCTTCTCCAGCAGCTATGCCTT CCTGCTGATTGCCAGTTCGCTGCAGGGCATCGGCTCGTCTGCTCCTCTGTGGCTGGG ATGGGCATGCTTGCCAGTGTCTACACAGATGATGAAGAGAGAGGCAACGTCATGGGAA TCGCCTTGGGAGGCTGGCCATGGGGGTCTTAGTGGGCCCCCTTCGGGAGTGTGCT CTATGAGTTTGTGGGAAGACGGCTCCGTTCTGCTGGTGTGCTGGCCGCTGGTACTCTTG GATGGAGCTATTAGCTCTTTGTGCTCCAGCGTCCCGGGTGCAGCCAGAGAGTCAGA AGGGGACACCCCTAACCACGCTGCTGAAGGACCCGTACATCCTCATTGCTGCAGGCTC CATCTGCTTTGCAAACATGGGCATCGCCATGCTGGAGCCAGCCCTGCCATCTGGATG ATGGAGACCATGTGTTCCGAAAGTGGCAGCTGGGCGTTGCCTTCTTGCCAGCTAGTA TCTCTTATCTCATTGGAACCAATATTTTGGGATACTTGACACAAAAATGGGGAGGTG GCTTTGTGCTCTTCTGGGAATGATAATTGTTGGAGTCAGCATTATGATTCCATT GCAAAAAACATTTATGGAATCATAGCTCCGAACCTTTGGAGTTGGTTTGAATTGGAA TGGTGGATTCTGTCATGATGCCTATCATGGGCTACCTCGTAGACCTGCGGCACGTGTC CGTCTATGGGAGTGTGTACGCCATTGCGGATGTGGCATTTGTATGGGGTATGCTATA GGTCTTCTGCTGGTGGTGTATTGCAAAGCAATTGGATTTCATGGCTCATGACAA TTATTGGGATAATTGATATTTCTTTTGGCCCTCTCTGCTTTTCTTCGAAGTCCACC TGCCAAAGAAGAAAAATGGCTATTCTCATGGATCACAACCTGCCCTATTAACAAAA ATGTACACTCAGAATAATATCCAGTCATATCCGATAGGTGAAGATGAAGAATCTGAAA GTGACTGAGATGAGATCCTCAAAAATCATCAAGTGTTAATTGTATAAAACAGTGTT TCCAGTGACACAACCTCATCCAGAACTGTCTTAGTCATACCATCCATCCCTGGTGAAG AGTAAAACCAAAGTTATTATTTCTTTCATGGTTATGGTTCGATTGCAACAGCCTT ATAAAGAAAAAGAGCTTTCTAGGGTTTGTATAAATAGTGTGAACTTTATTTTA TGTATTAAATTTTATTAAATATCATACAATATATTTTGTATGAAATAGGTATTGTGTAA ATCTATAAATATTGAATCCAAACCAATATAATTTCC		
	ORF Start: ATG at 40		ORF Stop: TGA at 1630
	SEQ ID NO: 338	530 aa	MW at 57130.4kD
NOV39b, CG158583-02 Protein Sequence	MALSELALVRWLQESRRSRKLILFIVFLALLLDNMLLTVVSSDPPPSASKAGITVP IIPSYLYSIKHEKNATEIQARPVHTASISDSFQSIFSYYDNSTMVTGNATRDLLHQT TATQHMVTNASAVPSDCPSEDKDLLNENVQVGLLFASKATVQLITNPFIGLLTNRIGY PIPIFAGFCIHVSTIMFAFSSSYAFLLIARSLQIGSSCSSVAGMGLASVYTDDEE RGNVMGIALGGLAMGVLVGPPFGSVLYEFVGKTAPFLVLAALVLLDGAIQFLVLPQSR VQPSQKGTPLTLLKDPYILIAAGSICFANMGIAMLEPALPIWMETMCSRKWQLGV AFLPASISYLIGTNIFGILAHKMGRWLCALLGMIIVGVSIILCIPFAKNIYGLIAPNFG VGFAIGMVDSSMMPIMGYLVDLRHVSVYGSVYAIADVAFCMGYATGPSAGGAIKAIG FPWLMTIIGIIDILFAPLCFFLRSPPAKEEKMAILMDHNCPIKTKMYTQNNIQSYPIG EDEESES		
	SEQ ID NO: 339	1647 bp	

NOV39c, CG158583-04 DNA Sequence	GGAGCCATGGCCCTGAGCGAGCTGGCGCTGGTCCGCTGGCTGCAGGAGAGCCGCCGCT CGCGGAAGCTCATCCTGTTTCATCGTGTTCCTGGCGCTGCTGCTGGACAACATGCTGCT CACTGTCTGTTAGAGAGAGGGTTTCTCCATGTTGGCCAGCCTGGTCTCGAACTCCTG ACCTCAGGTGATCCACCTGCCTCAGCTTCCCAAAGTCTGGAATTACAGTCCCCATCA TCCCAAGTTATCTGTACAGCATTAAAGCATGAGAAGAATGCTACAGAAATCCAGACGGC CAGGCCAGTGCACACTGCCTCCATCTCAGACAGCTTCCAGAGCATCTTCTCCTATTAT GATAACTCGACTATGGTCACCGGGAATGCTACCAGAGACCTGACACTTCATCAGACCG CCACACAGCACATGGTGACCAACGCGTCCGCTGTTCTTCCGACTGTCCCAGTGAAGA CAAAGACCTCCTGAATGAAAACGTGCAAGTTGGTCTGTTGTTTGCCTCGAAAGCCACC GTCCAGTCTATCACCACCCCTTTCATAGGACTACTGACCAACAGAATTGGCTATCCAA TTCCCATATTTGCGGGATTCTGTCATCATGTTTGTCTCAACAATTATGTTTGCCTTCTC CAGCAGCTATGCCTTCCCTGCTGATTGCCAGGTCGCTGCAGGGCATCGGCTCGTCTGCTG TCCTCTGTGGCTGGGATGGGCATGCTTGCCAGTGTCTACACAGATGATGAAGAGAGAG GCAACGTCTATGGGAATCGCTTGGGAGGCTGGCCATGGGGGTCTTAGTGGGCCCCC CTTCGGGAGTGTGCTCTATGAGTTTGTGGGGAAGACGGCTCCGTTCTTGGTGTGCGC GCCCTGGTACTCTTGGATGGAGCTATTGAGCTCTTTGTGCTCCAGCCGTCGCCGGTGC AGCCAGAGAGTCAAGAAGGGGACACCCCTAACACGCTGCTGAAGGACCCGTACATCCT CATTGCTGCAGGCTCCATCTGCTTTGCAAACATGGGCATCGCCATGCTGGAGCCAGCC CTGCCATCTGGATGATGGAGACCATGTGTTCCCGAAAGTGGCAGCTGGGCGTTGCCT TCTTGCCAGCTAGTATCTCTTATCTCATTGGAACCAATATTTTGGGATACTTGACA CAAAATGGGGAGGTGGCTTTGTGCTCTTCTGGGAATGATAATTGTTGGAGTCAGCACT TTATGTATTCCATTTGCAAAAAACATTTATGGACTCATAGCTCCGAACCTTGGAGTTG GTTTGTGCAATTGGAATGGTGGATTCTGCAATGATGCCTATCATGGGCTACCTCGTAGA CCTGCGGCACGTGTCCGTCTATGGGAGTGTGTACGCCATTGCGGATGTGGCATTTTGT ATGGGGTATGCTATAGGTCTTCTGCTGGTGGTGTATTGCAAAGGCAATTGGATTTC CATGGCTCATGACAATTATTTGGGATAATTGATATCTTTTGGCCCTCTCTGCTTTT TCTTCGAAGTCCACCTACCAAAGAAGAAAAAATGGCTATTCTCATGGATCACAACTGC CCTATTAAACAAAAATGTACACTCAGAATAGTATCCAGTCAATCCGATAGGTGAAG ATGAAGAATCTGAAAGTGACTGA
	ORF Start: ATG at 7 ORF Stop: TGA at 1645
	SEQ ID NO: 340 546 aa MW at 58903.4kD
NOV39c, CG158583-04 Protein Sequence	MALSELALVRWLQESRRSRKLIILFIVFLALLLDNMLLTVVVERFLHVGQPGLELLTS GDPPASASQSPGIVPIIPSYLYSIKHEKNATEIQARPVHTASISDSFQSIIFSYYDN STMVTGNATRDLTQHQTATQHMVTNASAVPSDCPSEDKDLLNENVQVGLLFASKATVQ LITNPFIFGLLTNRIGYPIPIFAGFCIMFVSTIMFAFSSSYAFLLIARSLQIGSSCSS VAGMGMLASVYTDDEERGNVMGIALGGLAMGVLVGPPFGSVLYEFVVGKTAFFVLAAAL VLLDGAIQFLVLPQSRVQPSQKGTPLTLLKDPYILIAAGSICFANMGIAMLEPALP IWMETMCSRKWLQGVAFPLASISYLGITNIFGILAHKMRWLCALLGMIIVGVSTLC IPFAKNIYGLIAPNFGVGFAGIMVDSSMMPIMGYLVDLRHVSVYGSVYAIADVAFCMG YAIGPSAGGAIKAIGFPWLMTIIGIIDLFAPLCFRLRSPPTKEEKMAILMDHNCPI KTKMYTQNSIQSYPIGEDESESED
	SEQ ID NO: 341 1666 bp
NOV39d, CG158583-05 DNA Sequence	GCAGGCATCGCAAGCGACCCGAGCGGAGCCCCGGAGCCATGGCCCTGAGCGAGCTGG CGCTGGTCCGCTGGCTGCAGGAGAGCCCGCTCGCGGAAGCTCATCCTGTTTCATCGT GTTCTTGGCGCTGCTGCTGGACAACATGCTGCTCACTGCTGCTGGGTCAAGCGATCCT CCTTCTCAGCCTCAAAGGAGCTGGGATTACAGTCCCCATCATCCCAAGTTATCTGT ACAGCATTAAAGCATGAGAAGAATGCTACAGAAATCCAGACGGCCAGGCCAGTGACAC TGCTCCATCTCAGACAGCTTCCAGGGCATCTTCTCCTATTATGATAACTCGACTATG GTCACCGGGAATGCTACCAGAGACCTGACACTTCATCAGACCGCCACACAGCACATGG TGACCAACGCGTCCGCTGTTCTTCCGACTGTCCAGTGAAGACAAAGACCTCCTGAA TGAAAACGTGCAAGTTGGTCTGTTGTTTGCCTCGAAAGCCACCGTCCAGCTCATCACC AACCTTTCATAGGACTACTGACCAACAGAATTGGCTATCCAATTCCCATATTTGCGG GATTCTGCATCATGTTTGTCTCAACAATTATGTTTGCCTTCTCCAGCAGCTATGCCTT CCTGCTGATTGCCAGGTGCTGCAGGGCATCGGCTCGTCTGCTCCTCTGTGGCTGGG ATGGGCATGCTTGCCAGTGTCTACACAGATGATGAAGAGAGAGGCAACGTATGGGAA TCGCCTTGGGAGGCTGGCCATGGGGGTCTTAGTGGGCCCCCCTTCGGGAGTGTGCT CTATGAGTTTGTGGGGAAGACGGCTCCGTTCTTGGTGTGCTGGCTGCCCTGGTACTCTTG GATGGAGCTATTGAGCTCTTGTGCTCCAGCCGTCCCGGTGCAGCCAGAGAGTCAAG

	AGGGGACACCCCTAACCACGCTGCTGAAGGACCCGTACATCCTCATTGCTGCAGGCTC CATCTGCTTTTGCAAAACATGGGCATCGCCATGCTGGAGCCAGCCCTGCCCATCTGGATG ATGGAGACCATGTGTTCCCGAAAGTGGCAGCTGGGCGTTGCCTTCTTGCCAGCTAGTA TCTCTTATCTCATTGGAACCAATATTTTTGGGATACTTGACACAAAAATGGGGAGGTG GCTTTGTGCTCTTCTGGGAATGATAATTGTTGGAGTCAGCATTTTATGTATTCCATTT GCAAAAACATTTATGGACTCATAGCTCCGAACCTTTGGAGTTGGTTTTGCAATTGGAA TGGTGGATTGCTCAATGATGCCTATCATGGGCTACCTCGTAGACCTGCGGCACGTGTC CGTCTATGGGAGTGTGTACGCCATTGCGGATGTGGCATTGTTGTATGGGGTATGCTATA GGTCCTTCTGCTGGTGGTGTCTATTGCAAAGGCAATTGGATTTCCATGGCTCATGACAA TTATTGGGATAATTGATATTCTTTTTGCCCTCTCTGCTTTTTTCTTCGAAGTCCACC TGCCAAAGAAGAAAAAATGGCTATTCTCATGGATCACAACAGCCCTATTAACAAAAA ATGTACACTCAGAATAATATCCAGTCATATCCGATAGGTGAAGTGAAGAATCTGAAA GTGACTGAGATGAGATCCTCAAAAATCATCAAAGTGAAGGG		
	ORF Start: ATG at 40		ORF Stop: TGA at 1630
	SEQ ID NO: 342	530 aa	MW at 57142.5kD
NOV39d, CG158583-05 Protein Sequence	MALSELALVRWLQESRRSRKLILFIVFLALLLDNMLLTIVVGSSDPFASKGAGITVP IIPSYLYSIKHEKNATEIQTARPVHTASISDSFQGFISYDINSTMTVGNATRDLTLHQ TATQHMVTNASAVPSDCPSSEDKLLNENVQVGLLFASKATVQLITNPFIGLLTNRIQY PIPIFAGFCIMFVSTIMFAFSSSYAFLLIARSLQIGSSCSSVAGMGLASVYTDDEE RGNVMGIALGGLAMGVLVGPPFGSVLYEFVGKTAPFLVLAALVLLDGAIQFLVLPQSR VQPESQKGTPLTLLKDPYILIAAGSICFANMGIAMLEPALPIWMMETMCSRKWLGV AFLPASISYLGITNIFGILAHKMGRWLCALLGMIIVGVSIILCIPFAKNIYGLIAPNFG VGFAIGMVDSSMMPIMGYLVDLRHVSVYGSVYAIADVAFCMGYAIGPSAGGAIKAIG FPWLMTIIGIIDILFAPLCFFLRSPPAKEBKMAILMDHNCPIKTKMYTQNNIQSYPIG EDESES		
	SEQ ID NO: 343	1618 bp	
NOV39e, CG158583-03 DNA Sequence	GCAGGCATCGCAAGCGACCCCGAGCGGAGCCCGGAGCCATGGCCCTGAGCGAGCTGG CGCTGGTCCGCTGGCTGCAGGAGAGCCGCGCTCGCGGAAGCTCATCCTGTTTCATCGT GTTCTGCGCTGCTGCTGGACAACATGCTGCTCACTGTCGTGGTCCCCATCATCCCA AGTTATCTGTACAGCATTAAAGCATGAGAAGAATGCTACAGAAATCCAGACGGCCAGGC CAGTGCACTGCCTCCATCTCAGACAGCTTCCAGAGCATCTTCTCCTATTATGATAA CTCGACTATGGTCACCGGAATGCTACCAGAGACCTGACACTTCATCAGACCGCCACA CAGCACATGGTGACCAACGCGTCCGCTGTTTCCTTCCGACTGTCCCAGTGAAGAGAGCA ACCTCCTGAATGAAAACGTGCAAGTTGGTCTGTTGTTTGCCTCGAAAGCCACCGTCCA GCTCATCACCACCTTTTCATAGGACTACTGACCAACAGAAATGGCTATCCAATTCCC ATATTTGCGGGATTCTGCATCATGTTTGTCTCAACAATTATGTTTGCCTTCTCCAGCA GCTATGCCTTCTGCTGATTGCCAGGTCGCTGCAGGGCATCGGCTCGTCTGCTCCTC TGTGGCTGGGATGGGCATGCTTGCCAGTGTCTACACAGATGATGAAAGAGAGAGGCAAC GTCATGGGAATCGCCTTGGGAGGCTTGCCATGGGGGTCTTAGTGGGCCCCCCTTCG GGAGTGTGCTCTATGAGTTTGTGGGAAGACGGCTCCGTTCTGCTGCTGGCCGCCCT GGTACTCTTGATGGAGCTATTACAGCTCTTTGTGCTCCAGCGTCCCGGGTGCAGCCA GAGAGTCAGAAGGGGACACCCCTAACCACGCTGCTGAAGGACCCGTACATCCTCATTG CTGCAGGCTCCATCTGCTTTGCAACATGGGCATCGCCATGCTGGAGCCAGCCCTGCC CATCTGGATGATGGAGACCATGTGTTCCCGAAAGTGGCAGCTGGGCGTTGCCTTCTTG CCAGCTAGTATCTCTTATCTCATTGGAACCAATATTTTTGGGATACTTGACACAAAAA TGGGGAGGTGGCTTTTGCTCTTTCTGGGAATGATAATTGTTGGAGTCAGCACTTTATG TATTCATTGCAAAAACATTTATGGACTCATAGCTCCGAACCTTGGAGTTGGTTTT GCAATTGGAATGGTGGATTGCTCAATGATGCCTATCATGGGCTACCTCGTAGACCTGC GGCAGCTGTCCGCTATGGGAGTGTGTACGCCATTGCGGATGTGGCATTGTTGTATGGG GTATGCTATAGGTCCTTCTGCTGGTGGTGTCTATTGCAAAGGCAATTGGATTTCATGG CTCATGACAATTATTGGGATAATTGATATTCTTTTTGCCCTCTCTGCTTTTTTCTTC GAAGTCCACCTGCCAAAGAAGAAAAAATGGCTATTCTCATGGATCACAACAGCCCTAT TAAACAAAAATGTACACTCAGAATAGTATCCAGTCATATCCGATAGGTGAAGATGAA GAATCTGAAAGTGACTGAGATGAGATCCTCAAAAATCATCAAAGTGAAGGG		
	ORF Start: ATG at 40		ORF Stop: TGA at 1582
	SEQ ID NO: 344	514 aa	MW at 55672.9kD
NOV39e,	MALSELALVRWLQESRRSRKLILFIVFLALLLDNMLLTIVVPIIPSYLYSIKHEKNAT		

CG158583-03 Protein Sequence	EIQTARPVHTASISDSFSQISFSYYDNSTMVTGNATRDLTLLHQTATQHMTNASAVPSD CPSEDKDLLNENVQVGLLFASKATVQLITNPPFIGLLTNRIGYPIPIFAGFCIMFVSTI MFAFSSSYAFLLIARSLQIGSSCSSVAGMGLASVYTDDEERGNVMGIALGGLAMGV LVGPPFGSVLYEFVGKTAPFLVLAALVLLDGAIQFLVLPQSRVQPSQKGTPLTLLK DPYILIAAGSICFANMGIAMLEPALPIWMMETMCSRKWQLGVAFLPASISYLGITNIF GILAHKMGRWLCALLGMIIVGVSTLCIPFAKNYGLIAPNFGVGFAIGMVDSSMMPIM GYLVDLRHVSVYGSVYAIADVAFCMGYAIGPSAGGAIKAIGFPWLMTTIGIIDILFA PLCFFLRSPPAKEEKMAILMDHNCPIKTKMYTQNSIQSYPIGEDEESES
---------------------------------	---

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 39B.

Table 39B. Comparison of NOV39a against NOV39b through NOV39e.		
Protein Sequence	NOV39a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV39b	1..546 1..530	522/546 (95%) 523/546 (95%)
NOV39c	1..546 1..546	543/546 (99%) 544/546 (99%)
NOV39d	1..546 1..530	523/546 (95%) 524/546 (95%)
NOV39e	1..546 1..514	512/546 (93%) 513/546 (93%)

Further analysis of the NOV39a protein yielded the following properties shown in Table 39C.

Table 39C. Protein Sequence Properties NOV39a	
PSort analysis:	0.6400 probability located in plasma membrane; 0.4600 probability located in Golgi body; 0.3700 probability located in endoplasmic reticulum (membrane); 0.1000 probability located in endoplasmic reticulum (lumen)
SignalP analysis:	Cleavage site between residues 38 and 39

- 5 A search of the NOV39a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 39D.

Table 39D. Geneseq Results for NOV39a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV39a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
ABB09288	Human solute carrier familv 18	1..546	514/546 (94%)	0.0

	member 2 (SLC18A2) protein SEQ ID NO:3 - Homo sapiens, 514 aa. [WO200222652-A2, 21-MAR-2002]	1..514	514/546 (94%)	
AAW38286	Human synaptic vesicle amine transporter protein - Homo sapiens, 514 aa. [US5688936-A, 18-NOV-1997]	1..546 1..514	514/546 (94%) 514/546 (94%)	0.0
AAR47342	Mammalian synaptic vesicle amine transporter protein - Homo sapiens, 514 aa. [WO9325699-A, 23-DEC-1993]	1..546 1..514	514/546 (94%) 514/546 (94%)	0.0
AAW38285	Rat synaptic vesicle amine transporter protein - Rattus rattus, 515 aa. [US5688936-A, 18-NOV-1997]	1..546 1..515	470/551 (85%) 490/551 (88%)	0.0
AAR47335	Mammalian synaptic vesicle amine transporter protein - Rattus rattus, 515 aa. [WO9325699-A, 23-DEC-1993]	1..546 1..515	470/551 (85%) 490/551 (88%)	0.0

In a BLAST search of public sequence databases, the NOV39a protein was found to have homology to the proteins shown in the BLASTP data in Table 39E.

Table 39E. Public BLASTP Results for NOV39a				
Protein Accession Number	Protein/Organism/Length	NOV39a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q05940	Synaptic vesicle amine transporter (Monoamine transporter) (Vesicular amine transporter 2) (VAT2) - Homo sapiens (Human), 514 aa.	1..546 1..514	514/546 (94%) 514/546 (94%)	0.0
Q9H3P6	Synaptic vesicle monoamine transporter - Homo sapiens (Human), 522 aa.	4..546 12..522	511/543 (94%) 511/543 (94%)	0.0
S29810	monoamine transport protein - human, 514 aa.	1..546 1..514	510/546 (93%) 510/546 (93%)	0.0
Q27963	Synaptic vesicle amine transporter (Monoamine transporter) (Vesicular amine transporter 2) (VAT2) - Bos taurus (Bovine), 517 aa.	1..546 1..517	471/549 (85%) 492/549 (88%)	0.0
A46374	resernine-sensitive vesicular	1..546	472/551 (85%)	0.0

	monoamine transporter - rat, 515 aa.	1..515	492/551 (88%)	
--	--------------------------------------	--------	---------------	--

Pfam analysis predicts that the NOV39a protein contains the domains shown in the Table 39F.

Table 39F. Domain Analysis of NOV39a			
Pfam Domain	NOV39a Match Region	Identities/ Similarities for the Matched Region	Expect Value
sugar_tr	98..516	66/523 (13%) 268/523 (51%)	0.019

#### Example 40.

The NOV40 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 40A.

5

Table 40A. NOV40 Sequence Analysis			
	SEQ ID NO: 345	1096 bp	
NOV40a, CG158964-01 DNA Sequence	GCAACCGGGGCAGGCCGTGCCGGCTGAGGAGGTCCTGAGGCTACAGAGCTGCCGCGGC TGGCACACGAGCGCCTCGGCCTAACCAGAGTGTTCGCGGGGGCTGTGAGGGGAGGGCC CCGGGCGCCATTGCTGGCGGTGGGAGCGCCGCCCGGTCTCAGCCCGCCCTCGGCTGCT CTCCTCCTCCGGCTGGGAGGGCCGTAGCTCGGGGCGTCGCCAGCCCCGGCCCGGGC TCGAGAATCAAGGGCCTCGGCCCGCTCCCGCAGCTCAGTCCATCGCCCTTGCCGGGC AGCCCGGGCAGAGACCATGTTTGACAAGACGCGGCTGCCGTACGTGGCCCTCGATGTG CTCTGCGTGTGCTGGATTATTCTTGGAGAAACCTGTCTGTTTACTGTAACCTTTTG CACTCAAATTCCTTTATCAGGAATAACTACATAGCCACTATTTACAAAGCCATTGGAA CCTTTTATTTGGTGCAGCTGCTAGTCAGTCCCTGACTGACATTGCCAAGTATTCAAT AGGCAGACTGCGGCCTCACTTCTTGGATGTTTGTGATCCAGATTGGTCAAAAATCAAC TGCAGCGATGGTTACATTGAATACTACATATGTCGAGGGAATGCAGAAAGAGTTAAGG AAGGCAGTTGTCTTCTATTCAAGCCAGGATGAAGGGAGACTGGGCAGACTCTTACGCCCC ACACTGCAATTTGGTCTTGTGCGGTATCCATTTATGTGGGCCCTTCTCGAGTTTCTG ATTATAAACACCACTGGAGCGATGTGTGACTGGACTCATTACGGGAGCTCTGGTTGC AATATTAGTTGCTGTATGTATCGGATTTCTTCAAAGAAAGAACTTCTTTTAAAGAA AGAAAAGAGGAGGACTCTCATACAACTCTGCATGAAACACCAACAACCTGGGAATCACT ATCCGAGCAATCACCAGCCTTGAAAGGCAGCAGGTGCCAGGTGAAGCTGGCCTGTT TTCTAAAGGAAAATGATTGCCACAAGGCAAGAGGATGCATCTTCTTCCTGG		
	ORF Start: ATG at 344		ORF Stop: TGA at 1007
	SEQ ID NO: 346	221 aa	MW at 25083.4kD
NOV40a, CG158964-01 Protein Sequence	MCSACCWII LGETLSVYCNLLHSNSFIRNNYIATIYKAIGTFLFGAAASQSLTDIAKY SIGRLRPHFLDVCDPDWSKINCS DGYIEYYICRGAERVKEGRLSFYSGHSSFSMYCM LFVALYLQARMKGDWARLLRPTLQFGLVAVSIYVGLSRVSDYKHHWSVDLTGLIQGAL VAILVAVYVSDFFKERTSFKERKEEDSHTTLHETPTTGHNHYPSNHQP		
	SEQ ID NO: 347	1388 bp	
NOV40b, CG158964-02 DNA Sequence	CGGCCGCGTGCACGCAACCGGGGCAGGCCGTGCCGGCTGAGGAGGTCCTGAGGCTACA GAGCTGCCGCGGCTGGCACACGAGCGCCTCGGCACTAACCAGAGTGTTCGCGGGGGCTG TGAGGGGAGGGCCCCGGGCGCCATTGCTGGCGGTGGGAGCGCCGCCCGGTCTCAGCCC GCCCTCGGCTGCTCTCTCTCCGCTGGGAGGGGCCGTAGCTCGGGGCGGTGCCAG CCCCGGCCCGGCTCGAGAATCAAGGGCCTCGGCCCGCTCCCGCAGCTCAGTCCATC		

	<div>GCCCTTGCCGGGCAGCCCGGGCAGAGACCATGTTTGACAAGACGCGGCTGCCGTACGT GGCCCTCGATGTGCTCTGCGTGTGCTGGATTATCTTGAGAAAACCCGTCTGTTTA CTGTAACCTTTTGCACCTCAAATTCCTTTATCAGGAATAACTACATAGCCACTATTTAC AAAGCCATTGGAACCTTTTTATTTGGTGCAGCTGCTAGTCAGTCCCTGACTGACATTG CCAAGTATTC AATAGGCAGACTGCGGCCTCACTTCTTGGATGTTTGTGATCCAGATTG GTCAAAAATCAACTGCAGCGATGGTTACATTGAATACTACATATGTCGAGGGAATGCA GAAAGAGTTAAGGAAGGCAGGTTGTCCTTCTATTTCAGGCCACTCTTCGTTTTCATGT ACTGCATGCTGTTTGTGGCACTTTATCTTCAAGCCAGGATGAAGGGAGACTGGGCAAG ACTCTTACGCCCCACACTGCAATTTGGTCTTGTGCGGTATCCATTTATGTGGGCCTT TCTCGAGTTTCTGATTATAAACACCACTGGAGCGATGTGTTGACTGGACTCATTACAGG GAGCTCTGGTTGCAATATTAGTTGCTGTATATGTATCGGATTTCTTCAAAGAAAGAAC TTCTTTTAAAGAAAGAAAAGAGGAGGACTCTCATACAACTCTGCATGAAACACCAACA ACTGGGAATCACTATCCGAGCAATCACCAGCCTTGAAAGGCAGCAGGGTGCCCAGGTG AAGCTGGCCTGTTTTCTAAAGGAAAATGATTGCCACAAGGCAAGAGGATGCATCTTTC TTCCTGGTGTACAAGCCTTTAAAGACTTCTGCTGCTGCTATGCCTCTTGGATGCACAC TTTGTGTGTACATAGTTACCTTTAACTCAGTGGTTATCTAATAGCTCTAAACTCATT AAAAAACTCCAAGCCTTCCACCAAAACAGTGCCCCACCTGTATACATTTTATTAAAA AAATGTAATGCTTATGTATAAACATGTATGTAATATGCTTTCTATGAATGATGTTTGA TTTAAATATAATACATATTTAAATGTATGGGAGAACCACCAAAAAAAAAAAAAA</div>		
	ORF Start: ATG at 357		ORF Stop: TGA at 1020
	SEQ ID NO: 348	221 aa	MW at 25083.4kD
NOV40b, CG158964-02 Protein Sequence	MCSACWIIIGETLSVYCNLLHSNSFIRNNYIATYKAI GTFLFGAAAQS LLDIAKY SIGRLRPHFLDVCDPDWSKINCS DGYIEYYICRGNAERVKEGRLSFYSGHSSFSMYCM LFVALYLQARMKGDWARLLRPTLQFGLVAVSIYVGLSRVSDYKHHWS DVL TGLIQGAL VAILVAVYVSDFFKERTSFKERKEEDSHTTLHETPTTGNHYPSNHQP		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 40B.

Table 40B. Comparison of NOV40a against NOV40b.		
Protein Sequence	NOV40a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV40b	1..221 1..221	221/221 (100%) 221/221 (100%)

Further analysis of the NOV40a protein yielded the following properties shown in Table 40C.

Table 40C. Protein Sequence Properties NOV40a	
PSort analysis:	0.6400 probability located in endoplasmic reticulum (membrane); 0.4960 probability located in plasma membrane; 0.3776 probability located in microbody (peroxisome); 0.1900 probability located in Golgi body
SignalP analysis:	Cleavage site between residues 49 and 50

A search of the NOV40a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 40D.

<b>Table 40D. Geneseq Results for NOV40a</b>				
<b>Geneseq Identifier</b>	<b>Protein/Organism/Length [Patent #, Date]</b>	<b>NOV40a Residues/ Match Residues</b>	<b>Identities/ Similarities for the Matched Region</b>	<b>Expect Value</b>
AAY24916	Human phosphatase HPA-1 - Homo sapiens, 285. aa. [WO9931225-A2, 24-JUN-1999]	8..221 72..285	214/214 (100%) 214/214 (100%)	e-125
AAW79284	Human phosphatidic acid phosphatase alpha 1 - Homo sapiens, 284 aa. [WO9846730-A1, 22-OCT-1998]	8..221 71..284	214/214 (100%) 214/214 (100%)	e-125
AAW79285	Human phosphatidic acid phosphatase alpha 2 - Homo sapiens, 285. aa. [WO9846730-A1, 22-OCT-1998]	8..221 72..285	213/214 (99%) 213/214 (99%)	e-124
AAW79287	Human phosphatidic acid phosphatase gamma - Homo sapiens, 276 aa. [WO9846730-A1, 22-OCT-1998]	11..200 72..260	123/190 (64%) 145/190 (75%)	2e-66
AAW79286	Human phosphatidic acid phosphatase beta - Homo sapiens, 311 aa. [WO9846730-A1, 22- OCT-1998]	8..192 100..283	113/185 (61%) 138/185 (74%)	5e-59

In a BLAST search of public sequence databases, the NOV40a protein was found to have homology to the proteins shown in the BLASTP data in Table 40E.

<b>Table 40E. Public BLASTP Results for NOV40a</b>				
<b>Protein Accession Number</b>	<b>Protein/Organism/Length</b>	<b>NOV40a Residues/ Match Residues</b>	<b>Identities/ Similarities for the Matched Portion</b>	<b>Expect Value</b>
O14494	PHOSPHATIDIC acid phosphatase 2A (EC 3.1.3.4) - Homo sapiens (Human), 284 aa.	8..221 71..284	214/214 (100%) 214/214 (100%)	e-124
O60463	Type-2 phosphatidic acid phosphohydrolase - Homo sapiens (Human), 289 aa.	8..221 76..289	214/214 (100%) 214/214 (100%)	e-124
O60457	Type-2 phosphatidic acid phosphatase alpha-2 (EC 3.1.3.4) - Homo sapiens (Human), 285 aa.	8..221 72..285	213/214 (99%) 213/214 (99%)	e-123
O88957	Phosphatidic acid phosphatase 2a2	8..221	199/215 (92%)	e-116

	- <i>Cavia porcellus</i> (Guinea pig), 286 aa.	72..286	208/215 (96%)	
O88956	Phosphatidic acid phosphatase 2a - <i>Cavia porcellus</i> (Guinea pig), 285 aa.	8..221 71..285	198/215 (92%) 208/215 (96%)	e-116

PFam analysis predicts that the NOV40a protein contains the domains shown in the Table 40F.

Table 40F. Domain Analysis of NOV40a			
Pfam Domain	NOV40a Match Region	Identities/ Similarities for the Matched Region	Expect Value
PAP2	37..188	62/174 (36%) 133/174 (76%)	1.5e-50

#### Example 41.

The NOV41 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 41A.

5

Table 41A. NOV41 Sequence Analysis			
	SEQ ID NO: 349	1524 bp	
NOV41a, CG159084-01 DNA Sequence	AACCAGATGGGAAAAATGTCTATCTGTACTTTTCAATCAACTGAGAAAGACGAGAAAA AAGAAGCTCACACTGGGAATGTAATGACTAATTTATCAAATGCTTATGCTTCAGATTT GTTGCTATGCAAAGATGACAAAGACTTAACAAAATATTTTCTGCTGCAAGTGGTAGAT ATTCTTCTACAGTATGCAAAAAACACCTTTGATGGTAAGAGTGAAATATTGGACTTCC ATCATCCTCATCAACTACTTGAGGTTTGGTTGGGTTTACCTTAGAACTGCCTGACCA CCCTGAATCTCTGGAACAGTTACTTGCTGATTGCACAGATACCTTAAAAATACAGTGTT AAAACAGGTCATCCTCGCTATTTTAACCAGCTGTCCAGTGGGTAGATATGACTGGAC TTGCAGGGGAATGGTTGACAGCCACTGCAAATACCAACCTGTTTACATATGAAATAGC CCAGTTTTTACTGTCTATGGAGACAATCTTCTCAAGAAAATGTATGAAATTATTGGC TGGGGGAAGAAACAAGCAGATGGAATATTTTCACCTGGTGGCAGTATATCAAGCCTTT ATGGTATTTTAGTAGCTCACTATAAAACAATATCCAGAGATAAAAAACAAAAGGCATGAC TGCACTTCCATGCATTGTATTATTTGTTTCTGAGCAAGGTCATTACTCAATAAAAAATA GCTGCAACAATTTGGGTATGGAATTGATAATGTAATTGAAGTAAAGTGTGATGAAA GGGGAAAGATGATTCCAGCTGAGTTAGAGAAAAATATATTACAAGCTAAAAAAAAGG TCAAATCCATTCTGTGTCTGTGCCACAGCCGAAGCACAGTGTACGGAGCCTTCGAC CCTCTCCCTGACATCGCTGATATTTGTGAGAAGCACAACTCTGGATGCATGTGGATG CAGCTTGGGGAGGTGGACTGCTGCTATCCAGAACTATTCTTATAAACTCAGTGGTAT TGAAAGGGCCAAGTCTGTGACCTGGAATCCACACAACTAATGGGTGTCCCTCTTCAG TGCTCTGCTATCTTGATCCGGGAAAAAGGCCTTCTAGATGCATGTAATCAGATGCAAG CTGAATATCTTTCCAGTCAGGTAAACTCTACAATGTTGACTTTGACACGGCGGATAA AACTATTCAGTGTGGCCGACATGTTGATATCTTCAAGCAGTGGTTAATGTGGAAAGCA AAGGGAACCTTTGGCTTTGAGGAACAAATCAACAAATATATGGAACCTTGCAAAATACT TCTATAAGGTTTTAAAGAAAAAAGATAACTTTAAGCTTGTGTTTGTATGCAGAGCCTGA GTTCACTAATGTCTGCTTCTGGTATTTCCAGCAAGGCTTAAACATATTCCAAAGGT TTTGAAAGAGATCAAGAATCCGAAAGGTAGCTCCAAGATTAAAGCAGATGATGA TGGAAGGCACAATCATGATAAGCTACCAGCCATGTGGAGACAAAGTAAATATTTTGCG AATGGTTTTTTTCTAA		
	ORF Start: ATG at 7		ORF Stop: TAA at 1522

	SEQ ID NO: 350	505 aa	MW at 57169.9kD
NOV41a, CG159084-01 Protein Sequence	MGKMSICTFQSTKDEKKEAHTGNVMTNLSNAYASDLLLCKDDKDLTKYFLLQVVDIL LQYAKNTFDGKSEILDFFHHPQLLEGLVGFTLELPDHPESLEQLLADCTDTLKYSVKT GHPRYFNQLSSGLDMTGLAGEWLTATANTNLTFTYEIAPVFTVMETILLKKMYEIIIGWG KKQADGIFSPGGSISSLYGILVAHYKQYPEIKTKGMTALPCIVLFVSEQGHYSIKIAA TILGIGIDNVIEVKCDERGMIPAELEKNILQAKKKGQTPFCVCATAGSTVYGAFDPL PDIADICEKHKLWMHVDAAWGGGLLSRNYSYKLSGIERAKSVTWNPHKLMGVPLQCS AILLIREKGLLDACNQMQAEYLFQSGKLYNVDFDTADKTIQCGRHVDIFKQWLMWKAKG TLGFEEQINKYMELAKYFYKVLKKKDNFKLVFDAEPEFTNVCWFYFPARLKHIPKGF RDQELRKVAPKIKAQMMMEGTIMISYQPCGDKVNI LRMVFF		

Further analysis of the NOV41a protein yielded the following properties shown in Table 41B.

Table 41B. Protein Sequence Properties NOV41a	
PSort analysis:	0.5819 probability located in microbody (peroxisome); 0.1000 probability located in mitochondrial matrix space; 0.1000 probability located in lysosome (lumen); 0.0000 probability located in endoplasmic reticulum (membrane)
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV41a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 41C.

5

Table 41C. Geneseq Results for NOV41a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV41a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAY57064	Glutamate decarboxylase 67. (GAD-67) amino acid sequence - Homo sapiens, 594 aa. [WO9956763-A1, 11-NOV-1999]	14..503 80..569	319/490 (65%) 387/490 (78%)	0.0
AAR27221	Full length brain GAD - Homo sapiens, 594 aa. [WO9214485-A, 03-SEP-1992]	14..503 80..569	319/490 (65%) 387/490 (78%)	0.0
AAR27220	Brain GAD #2 - Mus musculus, 593 aa. [WO9214485-A, 03-SEP-1992]	27..503 92..568	317/477 (66%) 378/477 (78%)	0.0
AAB03072	Chimeric human GAD67/rat GAD65 glutamic acid decarboxylase, SEQ ID NO:4 - Chimeric - Homo sapiens, 594 aa. [US6060593-A, 09-MAY-2000]	14..503 80..569	310/490 (63%) 388/490 (78%)	0.0
AAY33656	Chimeric rat GAD65/human GAD67	14..503	310/490 (63%)	0.0

	fusion protein 2 - Synthetic, 594 aa. [US5968757-A, 19-OCT-1999]	80..569	388/490 (78%)	
--	---	---------	---------------	--

In a BLAST search of public sequence databases, the NOV41a protein was found to have homology to the proteins shown in the BLASTP data in Table 41D.

Table 41D. Public BLASTP Results for NOV41a				
Protein Accession Number	Protein/Organism/Length	NOV41a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q9YI58	Glutamate decarboxylase 67 - Gallus gallus (Chicken), 590 aa.	14..503 76..565	322/490 (65%) 388/490 (78%)	0.0
B41935	glutamate decarboxylase (EC 4.1.1.15) 1 - human, 594 aa.	14..503 80..569	319/490 (65%) 387/490 (78%)	0.0
Q99259	Glutamate decarboxylase, 67 kDa isoform (EC 4.1.1.15) (GAD-67) (67 kDa glutamic acid decarboxylase) - Homo sapiens (Human), 594 aa.	14..503 80..569	319/490 (65%) 387/490 (78%)	0.0
S48135	glutamate decarboxylase (EC 4.1.1.15) - human, 593 aa.	14..503 79..568	318/490 (64%) 387/490 (78%)	0.0
S51776	glutamate decarboxylase (EC 4.1.1.15) - human, 593 aa.	14..503 79..568	318/490 (64%) 387/490 (78%)	0.0

PFam analysis predicts that the NOV41a protein contains the domains shown in the Table 41E.

Table 41E. Domain Analysis of NOV41a			
Pfam Domain	NOV41a Match Region	Identities/ Similarities for the Matched Region	Expect Value
pyridoxal_deC	78..452	136/401 (34%) 322/401 (80%)	6.9e-154

## 5 Example 42.

The NOV42 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 42A.

Table 42A. NOV42 Sequence Analysis			
	SEQ ID NO: 351	2990 bp	
NOV42a, CG159130-01	CCGGCGCCGGGCGGCGGCGAGTCTGGAGCCCGCGCCGTCGCCGGCCGCGTCTCTCCGG GCATGGAAGGAGGCGGAAGCCCAACTCTTCGTCTAACAGCCGGGACGATGGCAACAG		

DNA Sequence	CGTCTTCCCCGCCAAGGCGTCCGCGCCGGGCGCGGGCCGGCCGGCCGAGAACGCGC CTGGGCACCCCGCCGGGGGGCGGCGGGGCCGGCGCGAAGGAGCACGGCAACTCCGTGT GCTTCAAGGTGGACGGCGGTGGCGGGCGGTGGCGGGCGCGCGCGCGCGGAGAGCC GGCGGGGGGGCTTCAAGACGCCGAGGGGCCCCGGCGGCAGTACGGCTTCATGCAGAGG CAGTTCACCTCCATGCTGCAGCCCGGGGTCAACAAATCTCCCTCCGCATGTTTGGGA GCCAGAAGGCGGTGGAAAAGGAGCAGGAAAGGGTTAAAACTGCAGGCTTCTGGATTAT CCACCCTTACAGTGATTTTCAAGTGGGTTTAAATAATGCTTATAATGATGGTT GGAAATCTAGTCATCATAACAGTTGGAATCACATTCTTTACAGAGCAAACAACAC CATGGATTATTTTCAATGTGGCATCAGATACAGTTTTCCTATTGGACCTGATCATGAA TTTTAGGACTGGGACTGTCAATGAAGACAGTTCTGAAATCATCCTGGACCCCAAAGTG ATCAAGATGAATTATTTAAAAAGCTGGTTTGTGGTTGACTTCATCTCATCCATCCAG TGGATTATATCTTTCTTATTGTAGAAAAGGAATGGATTCTGAAGTTTACAAGACAGC CAGGGCACTTCGCATTGTGAGGTTTACAAAAATTCTCAGTCTCTTGGCTTTATTACGA CTTTCAAGGTTAATTAGATACATACATCAATGGGAAGAGATATTCACATGACATATG ATCTCGCCAGTGCAGTGGTGAGAAATTTTAAATCTCATCGGCATGATGCTGCTCCTGTG CCACTGGGATGGTTGTCTTCAGTTCCTAGTACCCTACTGCAGGACTTCCACCAGAT TGCTGGGTGTCTTTAAATGAAATGGTTAATGATTCTTGGGGAAAGCAGTATTCATACG CACTCTTCAAAGCTATGAGTCACATGCTGTGCATTGGGTATGGAGCCCAAGCCCCAGT CAGCATGTCTGACCTCTGGATTACCATGCTGAGCATGATCGTCGGGGCCACCTGCTAT GCCATGTTTGTGCGCCATGCCACCGCTTAAATCCAGTCTCTGGATTCTTCGAGGCGGC AGTATCAAGAGAAGTATAAGCAAGTGGAAACAATACATGTCATTCCATAAGTTACCAGC TGATATGCGTCAGAAGATACATGATTACTATGAACACAGATACCAAGGCCAAATCTTT GATGAGGAAAATATTCTCAATGAATCAATGATCCTCTGAGAGAGGAGATAGTCAACT TCAACTGTGCGAAACTGGTGGCTACAATGCCTTTATTGTCTAATGCGGATCCTAATTT TGTGACTGCCATGCTGAGCAAGTTGAGATTGAGGTGTTTCAACCTGGAGATTATATC ATACGAGAAGGAGCCGTGGGTAAAAAATGTATTTCAATCAACACGGTGTGCTGGTG TCATTACAAAATCCAGTAAAGAAATGAAGCTGACAGATGGCTCTTACTTTGGGAGAT TTGCCTGCTGACCAAAGGACGTCGTACTGCCAGTGTTCGAGCTGATACATATTGTCGT CTTTACTCACTTTCGGTGACAATTTCAACGAGGTCTTGGAGGAATATCCAATGATGA GGAGAGCCTTTGAGACAGTTGCCATTGACCGACTAGATCGAATAGGAAAGAAAAATTC AATTCTTCTGCAAAAGTTCAGAAGGATCTGAACACTGGTGTTTTCAACAATCAGGAG AACGAAATCCTCAAGCAGATTGTGAAACATGACAGGGAGATGGTGCAGGCAATCGCTC CCATCAATTATCCTCAAATGACAACCTGAATTCCACATCGTCTACTACGACCCCGAC CTCCCGCATGAGGACACAATCTCCACCGGTGTACACAGCGACCAGCCTGTCTCACAGC AACCTGCACTCCCCAGTCCCAGCACACAGACCCCCAGCCATCAGCCATCCTGTAC CCTGCTCCTACACCACCGCGGTCTGCAGCCCTCCTGTACAGAGCCCTTGCCCGCTCG AACTTCCACTATGCTTCCCCACCGCTCCCAGCTGTCACTCATGCAACAGCAGCGCG CAGCAGCAGGTACAGCAGTCCCAGCCGCGCAGACTCAGCCACAGCAGCCGTCCCCGC AGCCACAGACACCTGGCAGCTCCACGCCGAAAAATGAAGTGCACAAGAGCACGCAAGGC GCTTACAACACCAACCTGACCCGGGAAGTCAGGCCACTCTCCGCCTCGCAGCCCTCG CTGCCCCATGAGGTGTCACTCTGATTTCAGACCTCATCCACTGTGGGCGAGTCCC TGGCCTCCATCCCTCAACCCGTGACGGCGGTCCCCGGAACGGGCTTCAGGCAGGGG CAGGAGCACTGTCCCGCAGCGCGTCAACCCTCTTCCGACAGATGTCTGTCGGGAGCCATC CCCCCGAACCGAGGAGTCCCTCCAGCACCCCTCCACCAGCAGCTGCTCTTCCAAGAG AATCTTCTCAGTCTTAAACACAGACCCAGACGCAAGAAAGCCACGATTTGCTTCAAA TTTATGATCCCTGCTGATTGTCAAAGCAGAAAGAAATACTCTCATAACTGAGACTAT ACTCAGATCTTATTTTATCTATCTCTGATAGATCCCTCTAGCCTACTATGAAGAGA TATTTTAGACAGCTGTGGCTACACGTGAAATGTAAAAATATATATACATATACTATA AAATATATATCTAAATTCCCAAGAGAGGGTCAAAGACCTGTTTAGCATTCAGTGTTA TATGCTTTCCTTTCTTTAAATCATTAAAGGAT		
	ORF Start: ATG at 61		ORF Stop: TGA at 2731
	SEQ ID NO: 352	890 aa	MW at 98791.0kD
NOV42a, CG159130-01 Protein Sequence	MEGGGKPNSSNSRDDGNSVFPKASAPGAGPAAAEKRLGTPPGGGGAGAKEHGN SVC FKVDGGGGGGGGGGGEEFAGGFEDAEGRPRQYGFMRQFTSMLQPGVNFSLRMFGS QKAVEKEQERVKTAGFWIHPYSDFRYWDLIMLIMVGNLVIIPVGTFFTEQTTP WIFI NVASDTVFLDLIMNFRGTGTVNEDSSEIILDPKVIKMNLYKSWFVVDFISSIPV DYIFLIVEKGMDESVYKTARALRIVRFTKILSLRLRLRLIRYIHQWEEIFHMTYD LASAVVRIFNLI GMMLLLCHWDGCLQFLVPLLQDFPPDCWVSLNEMVNDSWGKQYSYA LFKAMSHMLCIGYGAQAPVMSDLWITMLSMIVGATCYAMFVGHATALIQSLDSSRRQ		

	YQEKYKQVEQYMSFHKLPADMRQKIHDYYEHRYQGKIFDEENILNELNDPLREEIVNF NCRKLVATMPLFANADPNFVTAMLSKLRFEVFPQGDYIIREGAVGKKMYFIQHGVAGV ITKSSKEMKLTGSGYFGEICLLTKGRRTASVRADTYCRLYSLSVDNFNEVLEEYPMR RAFETVAIDRLDRIGKNSILLQKFQKDLNTGVFNNQENEILKQIVKHDREMVQAIAP INYPQMTTLNSTSTTTPTSRMRTQSPPVYTATSLSHSNLHSPSPSTQTPQPSAILSP CSYTTAVCSPPVQSPLAARTFHYASPTASQLSLMQQPQQQVQSQPPQTQPPQPSPO PQTPGSSTPKNEVHKSTQALHNTNLTREVRPLSASQPSLPHEVSTLISRPHPTVGESL ASIPQPVTAVPGTGLQAGGRSTVPQRVTLFRQMSSGAIPPNRGVPPAPPPPPAAALPRE SSSVLNTDPDAEKPRFASNL
--	--

Further analysis of the NOV42a protein yielded the following properties shown in Table 42B.

Table 42B. Protein Sequence Properties NOV42a	
PSort analysis:	0.6000 probability located in plasma membrane; 0.4000 probability located in Golgi body; 0.3000 probability located in endoplasmic reticulum (membrane); 0.3000 probability located in microbody (peroxisome)
SignalP analysis:	No Known Signal Sequence Predicted

5 A search of the NOV42a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 42C.

Table 42C. Geneseq Results for NOV42a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV42a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAU11712	Human HCN1 channel subunit full length sequence from splice variant #1 - Homo sapiens, 890 aa. [WO200190142-A2, 29-NOV-2001]	1..890 1..890	890/890 (100%) 890/890 (100%)	0.0
AAU11714	Human full length HCN1 channel subunit variant 2 - Homo sapiens, 890 aa. [WO200190142-A2, 29-NOV-2001]	1..890 1..890	888/890 (99%) 888/890 (99%)	0.0
AAE18675	Human hyperpolarisation-activated cyclic nucleotide-gated channel 1 - Homo sapiens, 890 aa. [WO200202630-A2, 10-JAN-2002]	1..890 1..890	885/890 (99%) 885/890 (99%)	0.0
AAE21167	Human TRICH-11 protein - Homo sapiens, 882 aa. [WO200212340-A2, 14-FEB-2002]	1..890 1..882	882/890 (99%) 882/890 (99%)	0.0
AAY22191	Mouse brain CNG-1 protein	1..890	845/922 (91%)	0.0

	sequence - Mus sp, 910 aa. [WO9932615-A1, 01-JUL-1999]	1..910	852/922 (91%)	
--	---	--------	---------------	--

In a BLAST search of public sequence databases, the NOV42a protein was found to have homology to the proteins shown in the BLASTP data in Table 42D.

Table 42D. Public BLASTP Results for NOV42a				
Protein Accession Number	Protein/Organism/Length	NOV42a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
O88704	Hyperpolarization-activated cation channel, HAC2 - Mus musculus (Mouse), 910 aa.	1..890 1..910	846/922 (91%) 853/922 (91%)	0.0
Q9JKB0	Hyperpolarization-activated, cyclic nucleotide-gated potassium channel 1 - Rattus norvegicus (Rat), 910 aa.	1..890 1..910	847/922 (91%) 856/922 (91%)	0.0
O54899	Brain cyclic nucleotide gated 1 - Mus musculus (Mouse), 910 aa.	1..890 1..910	845/922 (91%) 852/922 (91%)	0.0
Q9MZS1	Hyperpolarization-activated cyclic nucleotide-gated channel 1 - Oryctolagus cuniculus (Rabbit), 822 aa.	78..890 14..822	786/813 (96%) 792/813 (96%)	0.0
O60741	Ion channel BCNG-1 - Homo sapiens (Human), 749 aa (fragment).	122..870 1..749	737/749 (98%) 739/749 (98%)	0.0

PFam analysis predicts that the NOV42a protein contains the domains shown in the Table 42E.

Table 42E. Domain Analysis of NOV42a			
Pfam Domain	NOV42a Match Region	Identities/ Similarities for the Matched Region	Expect Value
ion_trans	174..393	50/244 (20%) 160/244 (66%)	1.6e-22
cNMP_binding	490..578	31/120 (26%) 71/120 (59%)	2e-28
Transthyretin	692..709	12/19 (63%) 14/19 (74%)	0.82

## 5 Example 43.

The NOV43 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 43A.

Table 43A. NOV43 Sequence Analysis			
	SEQ ID NO: 353	1136 bp	
NOV43a, CG159178-01 DNA Sequence	AACACCATGAGGGCCCTGGTGCTTCTGCTGTCCCTGTTCTCTGCTGGGTGGCCAGGCCC AGCATGTGTCTGACTGGACCTACTCAGTGCAGATCGGCCTGCCCTCCACCATGCGCAT GACAGTGGCTGACGGCACTGTATACGTAGCCCAGCAGATGCACCTTCACTGGGGAGGT GCGTCTCGGAGATCAGCGGCTCTGAGCACACCGTGGACGGGATCAGACATGTGATCG AGATTCACATTGTTCACTACAATTCTAAATACAAGAGCTATGATATAGCCCAAGATGC GCCGATGGTTGGCTGTACTGGCAGCCTTCGTTGAGGTGAAGAATTACCTGAAAAC ACTTATTACAGCAACTTCATTTCTCATCTGGCCAACATCAAGTACCCAGGACAAAGAA CAACCCTGACTGGCCTTGACGTTCAAGGACATGCTGCCAGGAACCTCCAGCACTACTA CACCTACCATGGCTCACTCACCACGCCTCCCTGCACTGAGAACGTCCACTGGTTTGTG CTGGCAGATTTTGTCAAGCTCTCCAGGACACAGGTTTGAAGCTGGAGAATTCCTTAC TGGATCACCGCAATAAGACCATCCACAACGATTACCGCAGGACCCAGCCCCTGAAACA CAGAGTGGTGGAAATCCAACCTCCCGAATCAGGAATACACTCTAGGCTCTGAATTCAG TTTTACCTACATAAGATTGAGGAAATCTTGACTACTTAAGAAGAGCATTGAACGTAG GAAAGCTAAGAGGAAGATTCAATAATATTAAGTCTGAAGCCTGACCTAGCCAGAA GTGCTGTCCGCTGCAGCCGCACCCTACCTTGTCTAAGAAACCATGTGTGTCTGGAAC ACGCTGCTCCCTGGGCAGCTGTTGGGATCTGATTAAAGAGGGGAAACGATCATCCT GGACAGGAAGTGAATGGCTTCAGTTCATGAGACGGGATCTGAGTTAGACATCACCAG TGGAATTTGATTGGAATAGAACTTAAAGGAAATGGAACCTTAATATTCTCCCATCA AATCATATATGTGACCTGTCTGAATTATAAACAGCCTGACCTTCCTTTAGCATTA GATGTAATAAAATAACTTTTGAAATTTGTCATTT		
	ORF Start: ATG at 7		ORF Stop: TGA at 751
	SEQ ID NO: 354	248 aa	MW at 28657.2kD
NOV43a, CG159178-01 Protein Sequence	MRALVLLLSLFLGGQAQHVSDWTYSVQIGLPSTMRMTVADGTVYVAQQMHHFWGGAS SEISGSEHTVDGIRHVEIHIHVHNSKYKSYDIAQDAPDGLAVLAFVEVKNPENTY YSNFIHLNLIKYPGQRTTLTGLDVQDMLPRNLQHYTYHGLSTTPCTENVHWFVLA DFVKLSRTQVWKLNSLLDHRNKTIHNDYRRTQPLKRVVESNFPNQEYTLGSEFQFY LHKIEEILDYLRRLN		
	SEQ ID NO: 355	1006 bp	
NOV43b, CG159178-02 DNA Sequence	AACACCATGAGGGCCCTGGTGCTTCTGCTGTCCCTGTTCTCTGCTGGGTGGCCAGGCCC AGCATGTGTCTGACTGGACCTACTCAGAAGGGGCACTGGACGAAGCGCACTGGCCACA GCACTACCCCGCTGTGGGGGCCAGAGACAGTCGCCTATCAACCTACAGAGGACGAAG GTGCGGTACAACCCCTCCTGAAGGGGCTCAATATGACAGGCTATGAGACCCAGGCAG GGGAGTTCCCCATGGTCAACAATGGCCACACAGTGCAGATCGGCCTGCCCTCCACCAT GCGCATGACAGTGGCTGACGCACTGTATACATAGCCAGCAGATGCACCTTCACTGG GGAGGTGCGTCTCGGAGATCAGCGGCTCTGAGCACACCGTGGACGGGATCAGACATG TGATCGAGATTACATTGTTCACTACAATTCTAAATACAAGAGCTATGATATAGCCCA AGATGCGCCGGATGGTTTGGCTGTACTGGCAGCCTTCGTTGAGGTGAAGAATTACCT GAAAACACTTATTACAGCAACTTCATTTCTCATCTGGCCAACATCAAGTACCCAGGAC AAAGAACACCCCTGACTGGCCTTGACGTTCAAGGACATGCTGCCAGGAACCTCCAGCA CTACTACACCTACCATGGCTCACTCACCACGCCTCCCTGCACTGAGAACGTCCACTGG TTTGTGCTGGCAGATTTTGTCAAGCTCTCCAGGACACAGGTTTGAAGCTGGAGAATT CCTTACTGGATCACCGCAATAAGACCATCCACAACGATTACCGCAGGACCCAGCCCCT GAACCACAGAGTGGTGGAAATCCAACCTCCCGAATCAGGAATACACTCTAGGCTCTGAA TTCCAGTTTACCTACATAAGATTGAGGAAATCTTGACTACTTAAGAAGAGCATTGA ACTGAGGAAAGCTAAGAGGAAGATTCAATATTAAGTCTGAAGCCTGACCTAGCCA AGGGCGATTCCACACACTGC		
	ORF Start: ATG at 7		ORF Stop: TGA at 931
	SEQ ID NO: 356	308 aa	MW. at 35336.5kD
NOV43b, CG159178-02	MRALVLLLSLFLGGQAQHVSDWTYSEGALDEAHWPQHYPACGGQRQSPINLQRTKVR YNPSLKGLNMTGYETQAGEFPMVNNGHTVQIGLPSTMRMTVADGTVYIAQQMHHFWGG		

Protein Sequence	ASSEISGSEHTVDGIRHVIEIHI VHYSKYKSYDIAQDAPDGLAVLA AFVEVKNYPEN TYYSNFISHLANIKYPGQRTTLTGLDVQDMLPRNLQHYYTYHGSLLTPPCTENVHWFV LADFVKLSRTQVWKLNSLLDHRNKTIHNDYRRTQPLNHRVVESNFPNQEYTLGSEFQ FYLHKIEEILDYLRRLN
------------------	---

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 43B.

Table 43B. Comparison of NOV43a against NOV43b.		
Protein Sequence	NOV43a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV43b	25..248	220/224 (98%)
	85..308	223/224 (99%)

Further analysis of the NOV43a protein yielded the following properties shown in Table 43C.

Table 43C. Protein Sequence Properties NOV43a	
PSort analysis:	0.4132 probability located in outside; 0.2473 probability located in microbody (peroxisome); 0.1000 probability located in endoplasmic reticulum (membrane); 0.1000 probability located in endoplasmic reticulum (lumen)
SignalP analysis:	Cleavage site between residues 18 and 19

- 5 A search of the NOV43a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 43D.

Table 43D. Geneseq Results for NOV43a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV43a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAB59592	Human carbonic anhydrase isoform #5 - Homo sapiens, 262 aa. [US6160090-A, 12-DEC-2000]	25..219 68..262	189/195 (96%) 193/195 (98%)	e-112
AAE17175	Human RCC-associated antigen, G250 protein - Homo sapiens, 459 aa. [WO200198363-A2, 27-DEC-2001]	25..219 200..391	82/195 (42%) 112/195 (57%)	3e-37
AAB82848	Kidney cancer specific antigen G250-GM-CSF fusion protein - Homo sapiens, 610 aa. [WO200160317-A2, 23-AUG-2001]	25..219 345..536	82/195 (42%) 112/195 (57%)	3e-37

AAY53245	MN protein extracellular domain SEQ ID NO:87 - Homo sapiens, 377 aa. [US6027887-A, 22-FEB-2000]	25..219 163..354	82/195 (42%) 112/195 (57%)	3e-37
AAY53241	MN protein carbonic anhydrase domain SEQ ID NO:51 - Homo sapiens, 257 aa. [US6027887-A, 22- FEB-2000]	25..219 66..257	82/195 (42%) 112/195 (57%)	3e-37

In a BLAST search of public sequence databases, the NOV43a protein was found to have homology to the proteins shown in the BLASTP data in Table 43E.

Table 43E. Public BLASTP Results for NOV43a				
Protein Accession Number	Protein/Organism/Length	NOV43a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
P23280	Carbonic anhydrase VI precursor (EC 4.2.1.1) (Carbonate dehydratase VI) (CA-VI) (Secreted carbonic anhydrase) (Salivary carbonic anhydrase) - Homo sapiens (Human), 308 aa.	25..248 85..308	220/224 (98%) 224/224 (99%)	e-131
Q96QX8	DJ477M7.5 (carbonic anhydrase VI) - Homo sapiens (Human), 308 aa.	25..248 85..308	219/224 (97%) 222/224 (98%)	e-130
CRHU6	carbonate dehydratase (EC 4.2.1.1) VI precursor - human, 308 aa.	25..248 85..308	218/224 (97%) 222/224 (98%)	e-129
A29993	carbonate dehydratase (EC 4.2.1.1) VI - sheep, 307 aa.	25..245 68..291	164/224 (73%) 193/224 (85%)	1e-94
E966553	SYNTHETIC OVINE CARBONIC ANHYDRASE VI PROTEIN - vectors, 307 aa.	25..245 68..291	164/224 (73%) 193/224 (85%)	1e-94

Pfam analysis predicts that the NOV43a protein contains the domains shown in the Table 43F.

Table 43F. Domain Analysis of NOV43a			
Pfam Domain	NOV43a Match Region	Identities/ Similarities for the Matched Region	Expect Value
Carb_anhydrase	25..218	86/210 (41%) 191/210 (91%)	1.6e-118

#### 5 Example 44.

The NOV44 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 44A.

Table 44A. NOV44 Sequence Analysis			
	SEQ ID NO: 357	1704 bp	
NOV44a, CG160131-01 DNA Sequence	GGTTTCATGGCAGCCTCAAAGAAGGCAGTTTGGGGCCATTGGTGGGGCGGTGGACC AGGGCACCAGTTCGACGCGCTTTTGGTTTCAATTCAAAAACAGCTGAACACTTAG TCATCATCAAGTAGAAATAAAACAAGAGTTCCCAAGAGAAGGATGGGTGGAACAGGAC CCTAAGGAAATTCTACATTCTGTCTATGAGTGTATAGAGAAAACATGTGAGAACTTG GACAGCTCAATATTGATATTCCAACATAAAAGCTATTGGTGTGAGCAACCAGAGGGA AACCACTGTAGTCTGGGACAAGATAACTGGAGAGCCTCTTACAATGCTGTGGCTGCT CCAGTTTCTCCTGGCCCTTCAGTTCAGTTGCCGTTGTCCCTCTGGCTCTTCAGTTC CAGCTCCTGGTACTTCCTCAGTGTGGCTTGATCTAAGAACCCAGTCTACCGTTGAGAG TCTTAGTAAAAGAATTCCAGGAAATAATAACTTTGTCAAGTCCAAGACAGGCCCTTCCA CTTAGCACTTACTTCAGTGCAGTGAAACTTCGTTGGCTCCTTGACAAATGTGAGAAAAG TTCAAAAGGCCGTTGAAGAAAAACGAGCTCTTTTGGGACTATTGATTCATGGCTTAT TTGGAGTTTGACAGGAGGAGTCAATGGAGGTGTCCACTGTACAGATGTAACAAATGCA AGTAGGACTATGCTTTTCAACATTCATTCTTTGGAATGGGATAAACAACCTCTCGCAAT TTTTTGAATTCCAATGGAATTCTTCCAAATGTCCGGAGTTCCTCTGAGATCTATGG CCTAATGAAAGCTGGGGCCTTGAAGGTGTGCCAATATCTGGGTGTTTAGGGGACCAG TCTGCTGCATTGGTGGGACAAATGTGCTTCCAGATTGGACAAGCCAAAAATACGTATG GAACAGGATGTTTCTTACTATGTAATACAGGCCATAAGTGTGATTTTCTGATCATGG CCTTCTCACCACAGTGGCTTACAAACTTGGCAGAGACAAACCAGTATATATGCTTTG GAAGTTCTGTAGCTATAGCTGGTGTGTTATTCGCTGGCTAAGAGACAATCTTGAA TTATAAGACCTCAGAAGAAATTGAAAACTTGCTAAAGAAGTAGGTACTTCTTATGG CTGCTACTTCGTCCCAGCATTTTCGGGGTTATATGCACCTTATTGGGAGCCAGCGCA AGAGGGATAATCTGTGGACTCACTCAGTTCACCAATAAATGCCATATTGCTTTTGTCTG CATTAGAAGCTGTTTGTTCAAACTCGAGAGATTTTGGATGCCATGAATCGAGACTG TGGAATTCCTACTCAGTCATTTCAGGTAGATGGAGGAATGACCAGCAACAAATTCCTT ATGCAGCTACAAGCAGACATTCTGTATATACAGTAGTGAAGCCCTCAATGCCCGAAA CCACTGCACTGGGTGCGGCTATGGCGGCAGGGGCTGCAGAAGGAGTCGCGGTATGGAG TCTCGAACCCGAGGATTTGTCTGCCGTCACGATGGAGCGGTTTGAACCTCAGATTAAT GCGGAGGAAAGTGAATTCGTTATTCTACATGGAAGAAAGCTGTGATGAAGTCAATGG GTTGGGTTACAACCTCAATCTCCAGAAAGTGGTATTCCATAAAACCTACCAACTCATGG ATTCCCAAGATGTGAGCTTTTT		
	ORF Start: ATG at 7		ORF Stop: TAA at 1663
	SEQ ID NO: 358	552 aa	MW at 59929.2kD
NOV44a, CG160131-01 Protein Sequence	MAASKKAVLGPLVGAVDQGTSTRFLVFNSKTAELLSHHQVEIKQEFREGWVEQDPK EILHSVYECIEKTCCKLQNLNIDISNIKAIGVSNQRETTVVWDKITGEPLYNAAAPV SPGPSVPVAVVPSGSSVPAPGTSVSWLDLRTQSTVESLSKRIPGNFNVFKSKTGLPLS TYFSAVKLRWLLDNVRKVQKAVEEKRALFGTIDSWLIWSLTGGVNGGVHCTDVTNASR TMLFNIHSLEWDKQLCEFFGIPMEILPNVRSSEIYGLMKAGALEGVPI SGCLGDQSA ALVGQMCQFQIGQAKNTYGTGCFLLCNTGHKCVFSDHGLLTVAYKLGRDKPVYYALEG SVAIAGAVIRWLRDNLGIIKTSEEIEKLAKEVGTSGCYFVPFASGLYAPYWEPSARG IICGLTQFTNKCHIAFAALEAVCFQTREILDAMNRDCGIPLSHLQVDGGMSTNKKILMQ LQADILYIPVVKPSMPETALGAAMAAGAAEGVGVWSLEPEDLSAVTMERFEPQINAE ESEIRYSTWKKAVMKSMGWVTTQSPESGIP		
	SEQ ID NO: 359	1609 bp	
NOV44b, CG160131-04 DNA Sequence	CACCGGATCCATGGCAGCCTCAAAGAAGGCAGTTTGGGGCCATTGGTGGGGCGGTG GACCAGGGCACCAGTTCGACGCGCTTTTGGTTTCAATTCAAAAACAGCTGAACACT TTAGTCATCATCAAGTAGAAATAAAACAAGAGTTCCCAAGAGAAGGATGGGTGGAACA GGACCCTAAGGAAATTCTACATTCTGTCTATGAGTGTATAGAGAAAACATGTGAGAAA CTTGGACAGCTCAATATTGATATTCCAACATAAAAGCTATTGGTGTGAGCAACCAGA GGGAAACCACTGTAGTCTGGGACAAGATAACTGGAGAGCCTCTTACAATGCTGTGGT GTGGCTTGATCTAAGAACCCAGTCTACCGTTGAGAGTCTTAGTAAAAGAATTCCAGGA AATAATAACTTTGTCAAGTCCAAGACAGGCCCTTCCACTTAGCACTTACTTCAGTGCAG TGAAACTTCGTTGGCTCCTTGACAATGTGAGAAAAGTTCAAAGGCCGTTGAAGAAAA		

	ACGAGCTCTTTTTGGGACTATTGATTTCATGGCTTATTTGGAGTTTGACAGGAGGAGTC AATGGAGGTGTCCACTGTACAGATGTAACAAATGCAAGTAGGACTATGCTTTTTCAACA TTCATTCTTTGGAAATGGGATAAACAACCTGCGAATTTTTTGGAAATCCAATGGAAAT TCTTCCAAATGTCCGGAGTTCTTCTGAGATCTATGGCCTAATGAAAACTCTCATAGC GTGAAAGCTGGGGCCTTGGAAAGGTGTGCCAATATCTGGGTGTTTAGGGGACCAAGTCTG CTGCATTGGTGGGACAAATGTGCTTCCAGATTGGACAAGCCAAAAATACGTATGGAAC AGGATGTTTCTTACTATGTAATACAGGCCATAAGTGTGTATTTTCTGATCATGGCCTT CTCACCACAGTGGCTTACAACTTGGCAGAGACAAACCAGTATATTATGCTTTTGGAAAG GTTCTGTAGCTATAGCTGGTGTGTTATTCGCTGGCTAAGAGACAATCTTGGAAATTAT AAAGACCTCAGAAGAAATTGAAAACTTGCTAAAGAAGTAGGTACTTCTTATGGCTGC TACTTCGTCCCAGCATTTTTCGGGGTTATATGCACCTTATTTGGGAGCCAGCGCAAGAG GGATAATCTGTGGACTCACTCAGTTCACCAATAAATGCCATATTGCTTTTGTGCTGCATT AGAAGCTGTTTGTTCCAAACCTCGAGAGATTTTGGATGCCATGAATCGAGACTGTGGA ATTCCACTCAGTCATTTGCAGGTAGATGGAGGAATGACCAGCAACAAATCTTATGC AGCTACAAGCAGACATTCTGTATATACCAGTAGTGAAGCCCTCAATGCCCGAAACCAC TGCAGTGGGTGCGCTATGGCGGCAGGGGCTGCAGAAGGAGTCGGCGTATGGAGTCTC GAACCCGAGGATTGTCTGCCGTACGATGGAGCGGTTTGAACCTCAGATTAATGCGG AGGAAAGTGAAATTCGTTATTCTACATGGAAGAAAGCTGTGATGAAGTCAATGGGTG GGTTACAACCTCAATCTCAGAAAGTGGTATTCCAGTCGACGGC		
	ORF Start: at 2	ORF Stop: end of sequence	
	SEQ ID NO: 360	536 aa	MW at 58656.8kD
NOV44b, CG160131-04 Protein Sequence	TGSMAASKKAVLGLVGAVDQTSSTRFLVFNSKTAELLSHHQVEIKQEFREGWVEQ DPKEILHSVYECIEKTCEKLGQLNIDISNIKAIGVSNQRETTVVWDKITGEPLYNAV WDLRLTQSTVESLSKRI PGNNNFVSKTGLPLSTYFSAVKLRWLLDNVRKVQKAVEBK RALFGTIDSWLIWSLTGGVNGGVHCTDVTNASRTMLFNIHSELDWKQLEFFGIPMEI LPNVRSSSEIYGLMKISHSVKAGALEGVPI SGCLGDQSAALVQMCFQIGQAKNTYGT GCFLLCNTGHKCVFSDHGLLT VAYKLRDKPYYALEGSVAIAGAVIRWLRDNLGII KTSEEIEKLAKEVGTSYGCYFVPAFSGLYAPYWEPSARGIICGLTQFTNKCHIAFAAL EAVCFQTREILDAMNRDCGIPLSHLQVDGGMSTSNKILMQLQADILYI PVVKPSMPETT ALGAAMAAGAAEGVGVWSLEPEDLSAVTMERFEPQINAEBSEIRYSTWKAVMKSMGW VTTQSPESGIPVDG		
	SEQ ID NO: 361	1581 bp	
NOV44c, CG160131-02 DNA Sequence	GGTTTCATGGCAGCCTCAAAGAAGGCAGTTTTTGGGGCCATTGGTGGGGGCGGTGGACC AGGGCACCAGTTCGACGCGCTTTTGGTTTTCAATTCAAAAACAGCTGAACACTTAG TCATCATCAAGTAGAAAATAACAAGAGTTCCTCAAGAGAAGGATGGGTGGAACAGGAC CCTAAGGAAATTCATATTCTGTCTATGAGTGATAGAGAAAACATGTGAGAACTTG GACAGCTCAATATTGATATTTCCAACATAAAAGCTATTGGTGTGAGCAACCAGAGGGA AACCAGTGTAGTCTGGGACAAGATAACTGGAGAGCCTCTCTACAATGCTGTGGTGTGG CTTGATCTAAGAACCAGTCTACCGTTGAGAGTCTTAGTAAAGAATTCCAGGAAATA ATAACTTTGTCAAGTCCAAGACAGGCCTTCCACTTAGCACTTACTTCAGTGCAGTGAA ACTTCGTTGGCTCCTTGACAATGTGAGAAAAGTTCAAAGGCCGTGGAAGAAAACGA GCTCTTTTGGGACTATTGATTTCATGGCTTATTTGGAGTTTGACAGGAGGAGTCAATG GAGGTGTCCACTGTACAGATGTAACAAATGCAAGTAGGACTATGCTTTTCAACATTCA TTCTTTGGAAATGGGATAAACAACCTGCGAATTTTTTGGAAATCCAATGGAAATCTT CCAAATGTCCGGAGTTCTTCTGAGATCTATGGCCTAATGAAAGCTGGGGCCTTGGAA GTGTGCCAATATCTGGGTGTTTAGGGGACAGTCTGCTGCATTGGTGGGACAAATGTG CTTCAGATTGGACAAGCCAAAAATACGTATGGAACAGGATGTTTCTTACTATGTAAT ACAGGCCATAAGTGTGATTTTTCTGATCATGGCCTTCTCACCACAGTGGCTTACAAAC TTGGCAGAGACAAACCAGTATATTATGCTTTGGAAGGTTCTGTAGCTATAGCTGGTGC TGTTATTGCTGGCTAAGAGACAATCTTGGAAATATAAAGACCTCAGAAGAAATTGAA AAACTTGCTAAAGAAGTAGGTACTTCTTATGGCTGCTACTTCGTCCCAGCATTTTCGG GGTTATATGCACCTTATTGGGAGCCAGCGCAAGAGGGATAATCTGTGGACTCACTCA GTTACCAATAAATGCCATATTGCTTTTGTGTCATTAGAAGCTGTTTGTTCCTTCAAACT CGAGAGATTTGGATGCCATGAATCGAGACTGTGGAATTCCACTCAGTCATTGTCAGG TAGATGGAGGAATGACCAGCAACAAAATCTTATGCAGCTACAAGCAGACATTCTGTA TATACCAGTAGTGAAGCCCTCAATGCCCCGAAACCACTGCACTGGGTGCGGCTATGGCG GCAGGGGCTGCAGAAGGAGTCGGCGTATGGAGTCTCGAACCCGAGGATTGTCTGCCG		

	TCACGATGGAGCGGTTTGAACCTCAGATTAATGCCGAGGAAAGTGAAATTCGTTATTC TACATGGAAGAAAGCTGTGATGAAGTCAATGGGTGGGTACAACTCAATCTCCAGAA AGTGGTATTCCATAA		
	ORF Start: ATG at 7		ORF Stop: TAA at 1579
	SEQ ID NO: 362	524 aa	MW at 57488.5kD
NOV44c, CG160131-02 Protein Sequence	MAASKKAVLGPLVGAVDQGTSSSTRFLVFNSKTAELLSHHQVEIKQEFPREGWVEQDPK EILHSVYECIEKTCEKLGQLNIDISNIKAIGVSNQRETTVVWDKITGEPLYNAVWLD LRTQSTVESLSKRIPGNNNFVKSKTGLPLSTYFSAVKLRWLLDNVRKVQKAVEBKRAL FGTIDSWLIWLSLTGGVNGGVHCTDVTNASRTMLFNIHSLEWDKQLCEFFGI PMEILPN VRSSSEIYGLMKAGALEGVPI SGCLGDQSAALVGQMC FQIGQAKNTYGTGCFLLCNTG HKCVFSDHGLLT VAYKLGRDKPVYYALEGSVAIAGAVIRWLRDNLGIKTSEIEKL AKEVGTSGYCYFVPAPFSGLYAPYWEPSARGIICGLTQFTNKCHIAFAALEAVCFQTRE ILDAMNRDCGIPLSHLQVDGGMTSNKILMQLQADILYIPVVKPSPMPETTALGAAMAAG AAEGVGWVWSLEPEDLSAVTMERFEPQINAESEIRYSTWKKAVMKS MGWVTTQSPESG IP		
	SEQ ID NO: 363	1625 bp	
NOV44d, CG160131-03 DNA Sequence	TCGCCCTTTTGTACTGTATCGCCGAATTCATGGCAGCCTCAAAGAAGGCAGTTTGGG GCCATTGGTGGGGCGGTGGACCAGGCCACCAGTTCGACGCGCTTTTGGTTTTCAAT TCAAAAACAGCTGAAC TACTTAGTCATCATCAAGTAGAAAATAAACAGAGTTCCCAA GAGAAGGATGGGTGGAACAGGACCCTAAGGAAATTCATATTCTGTCTATGAGTGTAT AGAGAAAACATGTGAGAACTTGGACAGCTCAATATTGATATTTCCAACATAAAAGCT ATTGGTGTCAGCAACCAGAGGGAAACCACTGTAGTCTGGGACAAGATAACTGGAGAGC CTCTCTACAATGCTGTGGTGTGGCTTGATCTAAGAACCAGTCTACCGTTGAGAGTCT TAGTAAAAGAATTCCAGGAAATAATACTTTGTCAAGTCCAAGACAGGCCTTCCACTT AGCACTTACTTCAGTGCAGTGAAACTTCGTTGGCTCCTTGACAATGTGAGAAAAGTTC AAAAGGCCGTTGAAGAAAACAGAGCTCTTTTGGGACTATTGACTCATGGCTTATTG GAGTTTGACAGGAGGAGTCAATGGAGGTGTCCACTGTACAGATGTAACAATGCAAGT AGGACTATGCTTTTCAACATTCATTCTTTGGAATGGGATAAACAACCTCTGCGAATTTT TTGGAATTCGAATGGAATCTTCCAAATGTCCGGAGTTCTTCTGAGATCTATGGCCT AATGAAAGCTGGGGCCTTGAAGGTGTGCCAATATCTGGGTGTTTAGGGGACCAGTCT GCTGCATTGGTGGACAAAATGTGCTTCCAGATTGGACAAGCCAAAATACGTATGGAA CAGGATGTTTCTTACTATGTAATACAGGCCATAAGTGCCTATTTTCTGATCATGGCCT TCTCACCACAGTGGCTTACAACTTGGCAGAGACAAACCAGTATATTATGCTTTGGAA GGTCTGTAGCTATAGCTGGTGTGTTATTCGCTGGCTAAGAGACATCTTGGAAATTA TAAAGACCTCAGAAGAAATTGAAAACTTGCTAAAGAAGTAGGTACTTCTTATGGCTG CTACTTCGTCGCCAGCATTTTCGGGGTTATATGCACCTTATTGGGAGCCCAGCGCAAGA GGGATAATCTGTGACTCACTCAGTTCACCAATAAATGCCATATTGCTTTTGCTGCAT TAGAAGCTGTTTGTTCCTCAAACTCGAGAGATTTGGATGCCATGAATCGAGACTGTGG AATTCCACTCAGTCATTTGCAGGTAGATGGAGGAATGACCAGCAACAAAATCTTATG CAGCTACAAGCAGACATTCTGTATATACCAGTAGTGAAGCCCTCAATGCCCGAAACCA CTGCACTGGGTGCGGCTATGGCGGCAGGGGCTGCAGAAGGAGTCGGCGTATGGAGTCT CGAACCCGAGGATCTGTCTGCCGTACGATGGAGCGGTTGAACCTCAGATTAATGCG GAGGAAAGTGAAATTCGTTATTCTACATGGAAGAAAGCTGTGATGAAGTCAATGGGTT GGGTTACAACCTCAATCTCCAGAAAGTGGTATTCATGACTGCAGCCAACCTAATTCGG G		
	ORF Start: ATG at 30		ORF Stop: TGA at 1602
	SEQ ID NO: 364	524 aa	MW at 57502.5kD
NOV44d, CG160131-03 Protein Sequence	MAASKKAVLGPLVGAVDQATSSSTRFLVFNSKTAELLSHHQVEIKQEFPREGWVEQDPK EILHSVYECIEKTCEKLGQLNIDISNIKAIGVSNQRETTVVWDKITGEPLYNAVWLD LRTQSTVESLSKRIPGNNNFVKSKTGLPLSTYFSAVKLRWLLDNVRKVQKAVEBKRAL FGTIDSWLIWLSLTGGVNGGVHCTDVTNASRTMLFNIHSLEWDKQLCEFFGI PMEILPN VRSSSEIYGLMKAGALEGVPI SGCLGDQSAALVGQMC FQIGQAKNTYGTGCFLLCNTG HKCVFSDHGLLT VAYKLGRDKPVYYALEGSVAIAGAVIRWLRDNLGIKTSEIEKL AKEVGTSGYCYFVPAPFSGLYAPYWEPSARGIICGLTQFTNKCHIAFAALEAVCFQTRE ILDAMNRDCGIPLSHLQVDGGMTSNKILMQLQADILYIPVVKPSPMPETTALGAAMAAG AAEGVGWVWSLEPEDLSAVTMERFEPQINAESEIRYSTWKKAVMKS MGWVTTQSPESG IP		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 44B.

<b>Table 44B. Comparison of NOV44a against NOV44b through NOV44d.</b>		
<b>Protein Sequence</b>	<b>NOV44a Residues/ Match Residues</b>	<b>Identities/ Similarities for the Matched Region</b>
NOV44b	1..552	524/558 (93%)
	4..533	524/558 (93%)
NOV44c	1..552	524/552 (94%)
	1..524	524/552 (94%)
NOV44d	1..552	523/552 (94%)
	1..524	523/552 (94%)

Further analysis of the NOV44a protein yielded the following properties shown in Table 44C.

<b>Table 44C. Protein Sequence Properties NOV44a</b>	
<b>PSort analysis:</b>	0.4500 probability located in cytoplasm; 0.3731 probability located in microbody (peroxisome); 0.1000 probability located in mitochondrial matrix space; 0.1000 probability located in lysosome (lumen)
<b>SignalP analysis:</b>	No Known Signal Sequence Predicted

- 5 A search of the NOV44a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 44D.

<b>Table 44D. Geneseq Results for NOV44a</b>				
<b>Geneseq Identifier</b>	<b>Protein/Organism/Length [Patent #, Date]</b>	<b>NOV44a Residues/ Match Residues</b>	<b>Identities/ Similarities for the Matched Region</b>	<b>Expect Value</b>
ABB66928	Drosophila melanogaster polypeptide SEQ ID NO 27576 - Drosophila melanogaster, 538 aa. [WO200171042-A2, 27-SEP-2001]	10..548 17..529	277/542 (51%) 362/542 (66%)	e-155
AAU60271	Propionibacterium acnes immunogenic protein #21167. - Propionibacterium acnes, 526 aa. [WO200181581-A2, 01-NOV-2001]	15..542 28..520	266/530 (50%) 348/530 (65%)	e-144
ABB57950	Drosophila melanogaster polypeptide SEQ ID NO 642 -	12..545 32..537	251/538 (46%) 356/538 (65%)	e-143

	Drosophila melanogaster, 576 aa. [WO200171042-A2, 27-SEP-2001]			
ABB57948	Drosophila melanogaster polypeptide SEQ ID NO 636 - Drosophila melanogaster, 578 aa. [WO200171042-A2, 27-SEP-2001]	12..545 34..539	251/538 (46%) 356/538 (65%)	e-143
ABB57846	Drosophila melanogaster polypeptide SEQ ID NO 330 - Drosophila melanogaster, 576 aa. [WO200171042-A2, 27-SEP-2001]	12..545 32..537	251/538 (46%) 356/538 (65%)	e-143

In a BLAST search of public sequence databases, the NOV44a protein was found to have homology to the proteins shown in the BLASTP data in Table 44E.

Table 44E. Public BLASTP Results for NOV44a				
Protein Accession Number	Protein/Organism/Length	NOV44a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
P32189	Glycerol kinase (EC 2.7.1.30) (ATP:glycerol 3-phosphotransferase) (Glycerokinase) (GK) - Homo sapiens (Human), 524 aa.	1..552 1..524	524/552 (94%) 524/552 (94%)	0.0
Q14409	Glycerol kinase, testis specific 1 (EC 2.7.1.30) (ATP:glycerol 3-phosphotransferase) (Glycerokinase) (GK) - Homo sapiens (Human), 553 aa.	1..552 1..524	516/552 (93%) 518/552 (93%)	0.0
Q64516	Glycerol kinase (EC 2.7.1.30) (ATP:glycerol 3-phosphotransferase) (Glycerokinase) (GK) - Mus musculus (Mouse), 524 aa.	1..552 1..524	510/552 (92%) 521/552 (93%)	0.0
Q63060	Glycerol kinase (EC 2.7.1.30) (ATP:glycerol 3-phosphotransferase) (Glycerokinase) (GK) (ATP-stimulated glucocorticoid-receptor translocation promoter) (ASTP) - Rattus norvegicus (Rat), 524 aa.	1..552 1..524	510/552 (92%) 519/552 (93%)	0.0
Q14410	Glycerol kinase, testis specific 2 (EC 2.7.1.30) (ATP:glycerol 3-phosphotransferase) (Glycerokinase) (GK) - Homo sapiens (Human), 553 aa.	1..552 1..524	461/552 (83%) 495/552 (89%)	0.0

Pfam analysis predicts that the NOV44a protein contains the domains shown in the Table 44F.

Table 44F. Domain Analysis of NOV44a			
Pfam Domain	NOV44a Match Region	Identities/ Similarities for the Matched Region	Expect Value
FGGY	12..294	99/293 (34%) 266/293 (91%)	2.9e-126
FGGY_C	297..525	101/235 (43%) 222/235 (94%)	5.4e-110

#### Example 45.

5 The NOV45 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 45A.

Table 45A. NOV45 Sequence Analysis			
	SEQ ID NO: 365	1719 bp	
NOV45a, CG166282-01 DNA Sequence	GGCCGGACAGTCCGCCGAGGTGCTCGGTGGAGTCATGGCAGTGCCCTTTGTGGAAGAC TGGGACTTGGTGCAAACCTGGGAGAAGGTGCCATGGAGAAGTTCAACTTGCTGTGA ATAGAGTAACTGAAGAAGCAGTCGCAGTGAAGATTGTAGATATGAAGCGTGCCGTAGA CTGTCCAGAAAATATTAAGAAAAGAGATCTGTATCAATAAAATGCTAAATCATGAAAAT GTAGTAAAATTCTATGGTCACAGGAGAGAAGGCAATATCCAATATTTATTTCTGGAGT ACTGTAGTGGAGGAGAGCTTTTTGACAGAATAGAGCCAGACATAGGCATGCCTGAACC AGATGCTCAGAGATTCTTCCATCAACTCATGGCAGGGGTGGTTTATCTGCATGGTATT GGAATAACTCACAGGGATATTAACCAGAAAATCTTCTGTGGATGAAAGGGATAACC TCAAAATCTCAGACTTTGGCTTGGCAACAGTATTTTCGGTATAAATCGTGAGCGTTT GTTGAACAAGATGTGTGGTACTTTACCATATGTTGCTCCAGAACTTCTGAAGAGAAGA GAATTTTCATGCAGAACCAAGTTGATGTTTGGTCTGTGGAATAGTACTTACTGCAATGC TCGCTGGAGAATTGCCATGGGACCAACCCAGTGACAGCTGTCAGGAGTATTCTGACTG GAAAGAAAAAAAACATACCTCAACCCCTTGGAAAAAATCGATTCTGCCTCTAGCT CTGCTGCATAAAATCTTAGTTGAGAATCCATCAGCAAGAATTACCATTCCAGACATCA AAAAAGATAGATGGTACAACAAACCCCTCAAGAAAGGGGCAAAAAGGCCCGAGTCAC TTCAGGTGGTGTGTGAGTCTCCAGTGGATTTTCTAAGCACATTCAATCCAATTTG GACTTCTCTCCAGTAAACAGTGCTTCTAGTGAAGAAAATGTGAAGTACTCCAGTTCTC AGCCAGAACCCCGCACAGGTCTTTCCTTATGGGATACCAGCCCCCATACATTGATAA ATTGGTACAAGGGATCAGCTTTTCCAGCCACATGTCCTGATCATATGCTTTTGAAT AGTCAGTTACTTGGCACCCAGGATCCTCACAGAACCCCTGGCAGCGGTGGTCAAAA GAATGACACGATTTTATACCAATTGGATGCAGACAAATCTTATCAATGCCTGAAAGA GACTTGTGAGAAGTTGGGCTATCAATGGAAGAAAAGTTGTATGAATCAGGGTGATGGA TTGGAGTTCAAGAGACACTTCCTGAAGATTAAAGGGAAGCTGATTGATATTGTGAGCA GCCAGAAGGTTTGGCTTCCTGCCACATGATCGGACCATCGGCTCTGGGGAATCCTGGT GAATATAGTGTCTGATGTGACATTATTCTTCTAGAGAAGATTATCCTGTCTCTGCA AACTGCAATAGTAGTTCTCTGAAGTGTTCAC'TCCCTGTTTATCCAAACATCTTCCAA TTTATTTTGTGTTGTTCCGCATACAAATAATACCTATATCTTAATTGTAAGCAAACTT TGGGGAAGGATGAATAGAATTCAATTGATTATTTCTTCATGTGTGTTTAGTATCTGA ATTTGAAACTCATCTGTTGGAACCAAGTTTCAGGGGACATGAGTTTCCAGCTTTTA TACACACGTATCTCATTTTATCAAAACATTTTGT		
	ORF Start: ATG at 35		ORF Stop: TGA at 1361
	SEQ ID NO: 366	442 aa	MW at 50400.3kD

NOV45a, CG166282-01 Protein Sequence	MAVPFVEDWDLVQTLGEGAYGEVQLAVNRVTBEAVAVKIVDMKRAVDCPENIKKEICI NKMLNHENVVKFYGHRREGNIQYLFLEYCSGGELFDRIEPIGMEPEPDAQRFHQQLMA GVVYLHGIGITHRDIKPENLLDERDNLKISDFGLATVFRYNNRERLLNKMCGTLPYV APELLKRREFPHAEFVDVWSCGIVLTAMLAGELPWDQPSDSCQEYSDWKEKKTYLNPWK KIDSAPLALLHKILVENPSARITIPDIKKDRWYNKPLKKGAKRPRVTSGGVSESPSGF SKHIQSNLDFSPVNSASSEENVKYSSSQPEPTGLSLWDTSPSYIDKLVGGISFSQPT CPDHMLLNSQLLGTGSSQNPQRLVKRMTRFFTKLDADKSYQCLKETCEKLGQWKK SCMNQGDGLEFKRHFLKIKGKLIDIVSSQKVWLPAT
--	--

Further analysis of the NOV45a protein yielded the following properties shown in Table 45B.

Table 45B. Protein Sequence Properties NOV45a	
PSort analysis:	0.3000 probability located in nucleus; 0.1000 probability located in mitochondrial matrix space; 0.1000 probability located in lysosome (lumen); 0.0423 probability located in microbody (peroxisome)
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV45a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 45C.

5.

Table 45C. Geneseq Results for NOV45a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV45a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAU10752	Human checkpoint protein chk1 - Homo sapiens, 476 aa. [US6307015-B1, 23-OCT-2001]	1..442 1..476	442/476 (92%) 442/476 (92%)	0.0
AAE00662	Human cell cycle checkpoint protein, hchk1, alternative version #1 - Homo sapiens, 476 aa. [US6218109-B1, 17-APR-2001]	1..442 1..476	442/476 (92%) 442/476 (92%)	0.0
AAG68374	Human Chk1 kinase protein sequence - Homo sapiens, 476 aa. [WO200121771-A2, 29-MAR-2001]	1..442 1..476	442/476 (92%) 442/476 (92%)	0.0
AAE01155	Human Chk1 protein - Homo sapiens, 476 aa. [US6211164-B1, 03-APR-2001]	1..442 1..476	442/476 (92%) 442/476 (92%)	0.0
AAAY54452	A human checkpoint kinase (hChk1) polypeptide - Homo sapiens, 476 aa. [WO200003005-A2, 20-JAN-2000]	1..442 1..476	442/476 (92%) 442/476 (92%)	0.0

In a BLAST search of public sequence databases, the NOV45a protein was found to have homology to the proteins shown in the BLASTP data in Table 45D.

<b>Table 45D. Public BLASTP Results for NOV45a</b>				
<b>Protein Accession Number</b>	<b>Protein/Organism/Length</b>	<b>NOV45a Residues/ Match Residues</b>	<b>Identities/ Similarities for the Matched Portion</b>	<b>Expect Value</b>
O14757	Serine/threonine-protein kinase Chk1 (EC 2.7.1.-) - Homo sapiens (Human), 476 aa.	1..442 1..476	442/476 (92%) 442/476 (92%)	0.0
Q91ZN7	Checkpoint kinase 1 (Cell cycle checkpoint protein kinase) - Rattus norvegicus (Rat), 476 aa.	1..442 1..476	420/476 (88%) 430/476 (90%)	0.0
Q9D0N2	Checkpoint kinase 1 homolog (S. pombe) - Mus musculus (Mouse), 476 aa.	1..442 1..476	414/476 (86%) 428/476 (88%)	0.0
O35280	Serine/threonine-protein kinase Chk1 (EC 2.7.1.-) - Mus musculus (Mouse), 476 aa.	1..442 1..476	411/476 (86%) 427/476 (89%)	0.0
AAN33019	Checkpoint 1 protein - Gallus gallus (Chicken), 476 aa.	1..440 1..474	371/474 (78%) 403/474 (84%)	0.0

PFam analysis predicts that the NOV45a protein contains the domains shown in the Table 45E.

<b>Table 45E. Domain Analysis of NOV45a</b>			
<b>Pfam Domain</b>	<b>NOV45a Match Region</b>	<b>Identities/ Similarities for the Matched Region</b>	<b>Expect Value</b>
pkinase	9..265	93/294 (32%) 201/294 (68%)	1.2e-75

## 5. Example 46.

The NOV46 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 46A.

<b>Table 46A. NOV46 Sequence Analysis</b>			
	SEQ ID NO: 367	2264 bp	
NOV46a, CG170739-01. DNA Sequence	TTGACTGTATCGCCGGAATTCATGGCAGCGCCAGGCGGCAGGTCGGAGCCGCCGAGC TCCCCGAGTACAGCTGCAGCTACATGGTGTGCGCGCCGGTCTACAGCGAGCTCGCTTT CCAGCAACAGCACGAGCGGCGCCTGCAGGAGCGCAAGACGCTGCGGGAGAGCCTGGCC AAGTGCTGCAGTTGTTCAAGAAAGAGAGCCTTTGGTGTGCTAAAGACTCTAGTGCCCA		

	TCTTGGAGTGGCTCCCCAAATACCGAGTCAAGGAATGGCTGCTTAGTGACGTCATTTC GGGAGTTAGTACTGGGCTAGTGGCCACGCTGCAAGGACCTTTTCCAGTGGTGAGTTTA ATGGTGGGATCTGTTGTTCTGAGCATGGCCCCGACGAACACTTTCTCGTATCCAGCA GCAATGGAACGTGTATTAATACTACTATGATAGACACTGCAGCTAGAGATACAGCCAG AGTCTTGATTGCCAGTGCCCTGACTCTGCTGGTTGGAATTATACAGTTGATATTTGGT GGCTTGCAGATTGGATTTCATAGTGAGGCACTTGGCAGATCCTTTGGTTGGTGGCTTCA CAACAGCTGCTGCCTTCCAAGTGCTGGTCTCACAGCTAAAGATTGCTCTCAATGTTTC AACCAAAACTACAATGGAGTTCTCTCTATTATCTATACGCTGGTTGAGATTTTTCAA AATATGGTGATACCAATCTTGCTGATTTCACTGCTGGATTGCTCACCATTGTCGTCT GTATGGCAGTTAAGGAATTAATGATCGGTTTAGACACAAAATCCAGTCCCTATTCC TATAGAAGTAATTGTGACGATAATTGCTACTGCCATTTTCATATGGAGCCAACTGGAA AAAAATTACAAATGCTGGCATTTGTTAAATCCATCCCAAGGGGGTTTTGGCTCCTGAAC TTCCACCTGTGAGCTTGTTCTCGGAGATGCTGGCTGCATCATTTTCCATCGCTGTGGT GGCTTATGCTATTGCAGTGTGAGTAGGAAAAGTATATGCCACCAAGTATGATTACACC ATCGATGGGAACAGGAATTCATTGCCTTTGGGATCAGCAACATCTTCTCAGGATTCT TCTCTTGTGTTTGTGGCCACCACTGCTCTTTCCCGCACGGCCGTCCAGGAGAGCACTGG AGGAAAGACACAGGTTGCTGGCATCATCTCTGCTGCGATTGTGATGATCGCCATTCTT GCCCTGGGGAAGCTTCTGGAAACCCTTGCAGAAGTCGGTCTTGGCAGCTGTTGTAATTG CCAACCTGAAAGGGATGTTTATGCAGCTGTGTGACATTCTCGTCTGTGGAGACAGAA TAAGATTGATGCTGTTATCTGGGTGTTTACGTGTATAGTGTCCATATTCTGGGGCTG GATCTCGGTTTACTAGCTGGCCTTATATTTGGAAGTGTGACTGTGCTGCTGAGAGTTC AGTTTCCTTCTTGAATGGCCTTGGAAAGCATCCCTAGCACAGATATCTACAAAAGTAC CAAGAATTACAAAAACATTGAAGAACCCTCAAGGAGTGAAGATTCTTAGATTTTCCAGT CCTATTTTCTATGGCAATGTCGATGGTTTAAAAAATGTATCAAGTCCACAGTTGGAT TTGATGCCATTAGAGTATATAATAAGAGGCTGAAAGCGCTGAGGAAAAATACAGAACT AATAAAAAGTGGACAATTAGAGCAACGAAGAATGGCATCATAGTGATGCTGTTTCA ACAAATAATGCTTTTGAGCCCGATGAGGATATTGAAGATCTGGAGGAACCTTGATATCC CAACCAAGGAAATAGAGATTCAAGTGGATTGGAACCTGAGCTTCCAGTCAAAGTGAA CGTTCCCAAAGTGCCAATCCATAGCCTTGTGCTTGACTGTGGAGCTATATCTTTCCTG GACGTTGTTGGAGTGAGTCACTGCGGGTGATTGTCAAAGAATCCAAAGAATTGATG TGAATGTGTATTTTGCATCACTTCAAGATTATGTGATAGAAAAGCTGGAGCAATCGCG GTTCTTTGACGACAACATTAGAAAAGGACACATTCTTTTTCGCGTCCATGATGCTATA CTCTATCTACAGAACCAAGTGAAATCTCAAGAGGGTCAAGGTTCCATTTTAGAAACGA TCACTCTCATTAGGATTGTAAAGATACCCTTGAATTAGTAGAAACAGAGCTGACGGA AGAAGAAGTGTATGTCCAGGATGAGGCTATGCGTACACTTGCATCTGACTGCAGCCA AG		
	ORF Start: ATG at 22		ORF Stop: TGA at 2251.
	SEQ ID NO: 368	743 aa	MW at 81685.2kD
NOV46a, CG170739-01 Protein Sequence	MAAPGGRSEPPQLPEYSCSYMVSRPVYSELAFQQQHERRLQERKTLRESLAKCCSCSR KRAFGVLKTLVPILEWLPKYRVKEWLLSDVISGVSTGLVATLQGPFVPSLMVGSVVL SMAPDEHFLVSSSNGTVLNTTMDTAARDTARVLIASALTLLVGIIQLIFGGLQIGFI VRHLADPLVGGFTTAAAFQVLVSQKIVLNVSTKNYNGVLSIIYTLVEIFQNIQDNTNL ADFTAGLLTIIVCMVAVKELNDRFRHKIPVPIPIEVIVTIIATAISYGANLEKNYNAGI VKSIPRGFLPELPPVSLFSEMLAASFIAVVAYAIASVSGKVYATKYDYITIDGNQEF IAFGISNIFSGFFSCFVATTALSRATVQESTGGKTQVAGIISAAIVMIAILALGKLLLE PLQKSVLAADVIANLKGFMQLCDIPRLWRQNKIDAVIWWFTCVISIIILGLDLGLLAG LIFGLLTIVLRVQFPWNGLGSIPTSDIYKSTKNYKNIEEPQGVKILRFSSPIFYGNV DGFKKCIKSTVGFDVIRVYNKRLKALRKIQKLIKSGQLRATKNGIISDAVSTNNAFEP DEDIEDLEELDIPTKEIEIQVDWNSLPELVKVNVPKVIHSLVLDCCGISFLDVVGVR LRVIVKEFQRIDVNVYFASLQDYVIEKLEQCGFFDNRKDTFVLTVDAILYLQNV KSQEGQGSILETITLIQDKDTELELVETELTEELDVQDEAMRTLAS		

Further analysis of the NOV46a protein yielded the following properties shown in Table 46B.

**Table 46B. Protein Sequence Properties NOV46a**

PSort analysis:	0.8000 probability located in plasma membrane; 0.4000 probability located in Golgi body; 0.3000 probability located in endoplasmic reticulum (membrane); 0.0300 probability located in mitochondrial inner membrane
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV46a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 46C.

Table 46C. Geneseq Results for NOV46a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV46a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
ABG61914	Prostate cancer-associated protein #115 - Mammalia, 790 aa. [WO200230268-A2, 18-APR-2002]	1..743 1..780	741/780 (95%) 743/780 (95%)	0.0
AAM51696	Human pendrin SEQ ID NO 2 - Homo sapiens, 780 aa. [JP2001228146-A, 24-AUG-2001]	1..743 1..780	741/780 (95%) 743/780 (95%)	0.0
AAM51695	Mouse pendrin SEQ ID NO 1 - Mus sp, 780 aa. [JP2001228146-A, 24-AUG-2001]	1..743 1..780	648/780 (83%) 701/780 (89%)	0.0
AAR60568	Down-regulated in adenoma DRA tumor suppressor - Homo sapiens, 764 aa. [WO9420616-A, 15-SEP-1994]	20..692 9..720	322/716 (44%) 448/716 (61%)	e-176
AAG67162	Amino acid sequence of a human 32613 transporter polypeptide - Homo sapiens, 751 aa. [WO200164875-A2, 07-SEP-2001]	56..691 62..733	257/689 (37%) 401/689 (57%)	e-132

5. In a BLAST search of public sequence databases, the NOV46a protein was found to have homology to the proteins shown in the BLASTP data in Table 46D.

Table 46D. Public BLASTP Results for NOV46a				
Protein Accession Number	Protein/Organism/Length	NOV46a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
O43511	Pendrin (Sodium-independent chloride/iodide transporter) - Homo	1..743 1..780	741/780 (95%) 743/780 (95%)	0.0

	sapiens (Human), 780 aa.			
Q9R154	Pendrin (Sodium-independent chloride/iodide transporter) - Rattus norvegicus (Rat), 780 aa.	1..743 1..780	656/780 (84%) 700/780 (89%)	0.0
Q9R155	Pendrin (Sodium-independent chloride/iodide transporter) - Mus musculus (Mouse), 780 aa.	1..743 1..780	648/780 (83%) 701/780 (89%)	0.0
Q924C9	Chloride anion exchanger (DRA protein) (Down-regulated in adenoma) - Rattus norvegicus (Rat), 757 aa.	20..692 9..713	330/715 (46%) 470/715 (65%)	0.0
Q9WVC8	Chloride anion exchanger (DRA protein) (Down-regulated in adenoma) - Mus musculus (Mouse), 757 aa.	20..692 9..713	328/715 (45%) 463/715 (63%)	0.0

Pfam analysis predicts that the NOV46a protein contains the domains shown in the Table 46E.

Table 46E. Domain Analysis of NOV46a			
Pfam Domain	NOV46a Match Region	Identities/ Similarities for the Matched Region	Expect Value
7tm_3	171..410	48/293 (16%) 137/293 (47%)	0.46
Xan_ur_permease	85..465	67/468 (14%) 234/468 (50%)	0.56
Sulfate_transp	166..476	110/328 (34%) 265/328 (81%)	1.8e-97
STAS	499..688	32/192 (17%) 147/192 (77%)	1.6e-30

#### Example 47.

5 The NOV47 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 47A.

Table 47A. NOV47 Sequence Analysis			
	SEQ ID NO: 369	1337 bp	
NOV47a, CG171632-01 DNA Sequence	ATGAGATTTGGCATCTTTCTTTTGTGGTGGGGATGGGTTTGGCCACTGAAAGCAGAA TGCGCTGGCCCGGAAGAGAAGTCCACGAGATGTCTAAGAAAGGCAGTAGGCCCCAAAG ACAAAGACGAGAAGTACATGAAGATGCCACAAGCAAGTCAGCCCAATTCTGAGACGA AGTCCTGCCATTCTGTGGTGTGGATGTGCAGGTGGAGAGTTTGGATAGCATCTCAG AGGTTGACATGGACTTTACGATGACCCTCTACCTGAGGCACTACTGGAAGGACGAGAG		

	GCTGTCTTTTCCAAGCACCAACCACTCAGCATGACGTTTGATGGCCGGCTGGTCAAG AAGATCTGGGTCCCTGACATGTTTTCTGTCCTCCAAACGCTCCTTCATCCACGACA CCACCACAGACAACGTCATGTTGCGGGTCCAGCCTGATGGGAAAGTGCTCTATAGTCT CAGGGTTACAGTAAGTGAATGTGCAACATGGACTTCAGCCGATTTCCCTTGGACACA CAAACGTGCTCTCTTGAATTTGAAAGCTATGCCTATACAGAAGATGACCTCATGCTGT ACTGGAAAAGGGCAATGACTCCTTAAAGACAGATGAACGGATCTCACTCTCCAGTT CCTCATTCAGGAATTCACACCACCACCAAACTGGCTTTCTACAGCAGCACAGGCTGG TACAACCGTCTCTACATTAATTTACGTTGCGTGCACATCTTCTTCTTCTTGCCCC AAACTATTTCCCGCTACCCTGGTGGTCATGCTGTCTGGGTGTCCTTCTGGATCGA CCGAGAGCCGTGCTGCCAGAGTCCCCTTAGGTATCACAACGGTGCTGACCATGTCC ACCATCATCACGGGCGTGAATGCCTCCATGCCGCGCTCTCCTACATCAAGGCCGTGG ACATCTACCTCTGGGTGAGCTTTGTGTTCTGTTCTCTCGGTGCTGGAGTATGCGGC CGTCAACTACCTGACCACTGTGAGGAGAGGAAGGAACAGAAGCTGCGGGAGAAGCTT CCCTGCACAGCGGATTACCTCCGCCCAACACTGCGATGCTGGACGGCAACTACAGTG ATGGGGAGGTGAATGACCTGGACAACATACATGCCAGAGAATGGAGAGAAGCCCCGACAG GATGATGGTGCAGCTGACCTGGCCTCAGAGAGGAGCTCCCCACAGAGGAAAAGTCAG AGAAGCAGCTATGTGAGCATGAGAATCGACACCCACGCCATTGATAAATACTCCAGGA TCATCTTTCCAGCAGCATACATTTTATTCAATTTAATATACTGGTCTATTTTCTCCTA GAT		
	ORF Start: ATG at 1		ORF Stop: TAG at 1333
	SEQ ID NO: 370	444 aa	MW at 51932.2kD
NOV47a, CG171632-01 Protein Sequence	MRFGIFLLWGWVLATESRMRWPGREVHEMSKKGSRPQRQRREVHEDAHKQVSPILRR SPAIPVGVDVQVESLDSISEVDMDFMTLYLRHYWKDERLSFPSTNNLSMTFDGRILVK KIWVPDMFFVHSKRSFIHDTTNDNMLRVQPDGKVLVSLRVVTAMCNMDFSFPPLDT QTCSEIESYAYTEDDLMLYWKGNDSLKTDERISLSQFLIQEFHTTTKLAFYSSTGW YNRLYINFTLRRHIFFFLPQTYFPATLVVMSLWVSFWIDRRAPVAPRVPLGITTVLTMS TIITGVNASMPRVSYIKAVDIYLWVSFVFLSVLEYAAVNYLTTVQERKEQKLREKL PCTSGLPPPNNTAMLDGNYSDGEVNDLDNYMPENGEKPDMMVQLTLASERSSPQRKSQ RSSYVSMRIDTHAIDKYSRIIFPAAYILFNLIYWSIFS		
	SEQ ID NO: 371	1337 bp	
NOV47b, CG171632-01 DNA Sequence	ATGAGATTTGGCATCTTTCTTTTGTGGTGGGGATGGGTTTGGCCACTGAAAGCAGAA TGCGCTGGCCCGGAAGAGAAGTCCACGAGATGTCTAAGAAAGGCAGTAGGCCCAAAG ACAAAGACGAGAAGTACATGAAGATGCCCAAGCAAGTCAGCCCAATTCTGAGACGA AGTCTGCCATTCCTGTTGGTGTGGATGTGCAGGTGGAGAGTTTGGATAGCATCTCAG AGGTTGACATGGACTTTACGATGACCTCTACCTGAGGCCTACTGGAAGGACGAGAG GCTGTCTTTTCCAAGCACCAACCTCAGCATGACGTTTATGGCCGGCTGGTCAAG AAGATCTGGGTCCCTGACATGTTTTCGTGCACTCCAAACGCTCCTTCATCCACGACA CCACCACAGACAACGTCATGTTGCGGGTCCAGCCTGATGGGAAAGTGCTCTATAGTCT CAGGGTTACAGTAAGTGAATGTGCAACATGGACTTCAGCCGATTTCCCTTGGACACA CAAACGTGCTCTCTTGAATTTGAAAGCTATGCCTATACAGAAGATGACCTCATGCTGT ACTGGAAAAGGGCAATGACTCCTTAAAGACAGATGAACGGATCTCACTCTCCAGTT CCTCATTAGGAATTCACACCACCAACCAACTGGCTTTCTACAGCAGCACAGGCTGG TACAACCGTCTCTACATTAATTTACGTTGCGTGCACATCTTCTTCTTCTTGCCCC AAACTATTTCCCGCTACCCTGGTGGTCATGCTGTCTGGGTGTCCTTCTGGATCGA CCGAGAGCCGTGCTGCCAGAGTCCCCTTAGGTATCACAACGGTGCTGACCATGTCC ACCATCATCACGGGCGTGAATGCCTCCATGCCGCGCTCTCCTACATCAAGGCCGTGG ACATCTACCTCTGGGTGAGCTTTGTGTTCTGTTCTCTCGGTGCTGGAGTATGCGGC CGTCAACTACCTGACCACTGTGAGGAGAGGAAGGAACAGAAGCTGCGGGAGAAGCTT CCCTGCACAGCGGATTACCTCCGCCCAACACTGCGATGCTGGACGGCAACTACAGTG ATGGGGAGGTGAATGACCTGGACAACATACATGCCAGAGAATGGAGAGAAGCCCCGACAG GATGATGGTGCAGCTGACCTGGCCTCAGAGAGGAGCTCCCCACAGAGGAAAAGTCAG AGAAGCAGCTATGTGAGCATGAGAATCGACACCCACGCCATTGATAAATACTCCAGGA TCATCTTTCCAGCAGCATACATTTTATTCAATTTAATATACTGGTCTATTTTCTCCTA GAT		
	ORF Start: ATG at 1		ORF Stop: TAG at 1333
	SEQ ID NO: 372	444 aa	MW at 51932.2kD
NOV47b,	MRFGIFLLWGWVLATESRMRWPGREVHEMSKKGSRPQRQRREVHEDAHKQVSPILRR		

CG171632-01 Protein Sequence	SPAIPVGVDVQVESLDSISEVDMDFMTLTLYLRHYWKDERLSFPSTNNLSMTFDGRLVK KIWVPDMFFVHRSKRSFIHDTTDDNVMLRVQPDGKVLVSLRVTVTAMCNMDFSRLFPLDT QTCSEIESYAYTEDDLMLYWKKGNDLSLKTDERISLSQFLIQEFHTTTKLAFYSSTGW YNRLYINFTLRRHIFFFLPQTYFPATLVVMLSWSVFWIDRRVAPARVPLGITTVLTMS TTITGVNASMPRVSYIKAVDIYLWVSFVFVFLSVLEYAANVYLTTVQERKEQKLREKL PCTSGLPPEPNTAMLDGNYSDGEVNDLDNYPENGEKPDMMVQLTLASERSSPQRKSQ RSSYVSMRIDTHAIDKYSRIIFPAAYILFNLIYWSIFS
---------------------------------	--

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 47B.

Table 47B. Comparison of NOV47a against NOV47b.		
Protein Sequence	NOV47a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV47b	1..444	444/444 (100%)
	1..444	444/444 (100%)

Further analysis of the NOV47a protein yielded the following properties shown in Table 47C.

Table 47C. Protein Sequence Properties NOV47a	
PSort analysis:	0.4600 probability located in plasma membrane; 0.1692 probability located in microbody (peroxisome); 0.1000 probability located in endoplasmic reticulum (membrane); 0.1000 probability located in endoplasmic reticulum (lumen)
SignalP analysis:	Cleavage site between residues 16 and 17

- 5 A search of the NOV47a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 47D.

Table 47D. Geneseq Results for NOV47a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV47a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAE21956	Human transporter protein - Homo sapiens, 467 aa. [US2002028773-A1, 07-MAR-2002]	18..443 36..466	268/432 (62%) 320/432 (74%)	e-149
AAU04467	Human gamma-amino butyric acid (GABA) receptor protein #1 - Homo sapiens, 467 aa. [WO200153489-A1, 26-JUL-2001]	18..443 36..466	268/432 (62%) 320/432 (74%)	e-149
AAU04470	Human gamma-amino butyric acid	35..443	263/412 (63%)	e-149

	(GABA) receptor protein #4 - Homo sapiens, 420 aa. [WO200153489-A1, 26-JUL-2001]	9..419	313/412 (75%)	
AAG68256	Human POLY3 protein sequence SEQ ID NO:6 - Homo sapiens, 468 aa. [WO200179294-A2, 25-OCT-2001]	18..443 36..467	266/433 (61%) 318/433 (73%)	e-146
AAO14188	Human transporter and ion channel TRICH-5 - Homo sapiens, 467 aa. [WO200204520-A2, 17-JAN-2002]	18..443 36..466	264/432 (61%) 317/432 (73%)	e-146

In a BLAST search of public sequence databases, the NOV47a protein was found to have homology to the proteins shown in the BLASTP data in Table 47E.

Table 47E. Public BLASTP Results for NOV47a				
Protein Accession Number	Protein/Organism/Length	NOV47a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
P24046	Gamma-aminobutyric-acid receptor rho-1 subunit precursor (GABA(A) receptor) - Homo sapiens (Human), 473 aa.	1..444 1..473	439/474 (92%) 440/474 (92%)	0.0
P50572	Gamma-aminobutyric-acid receptor rho-1 subunit precursor (GABA(A) receptor) - Rattus norvegicus (Rat), 474 aa.	1..444 1..474	416/474 (87%) 425/474 (88%)	0.0
P56475	Gamma-aminobutyric-acid receptor rho-1 subunit precursor (GABA(A) receptor) - Mus musculus (Mouse), 474 aa.	1..444 1..474	413/474 (87%) 423/474 (89%)	0.0
Q8UW04	GABA receptor rho-1 subunit - Fugu rubripes (Japanese pufferfish) (Takifugu rubripes), 480 aa.	23..443 54..479	325/427 (76%) 361/427 (84%)	0.0
Q9YGQ4	Gamma-aminobutyric-acid receptor rho-1A subunit - Morone americana (White perch), 476 aa.	60..444 89..476	317/389 (81%) 345/389 (88%)	0.0

PFam analysis predicts that the NOV47a protein contains the domains shown in the Table 47F.

Table 47F. Domain Analysis of NOV47a			
Pfam Domain	NOV47a Match Region	Identities/	Expect Value

		Similarities for the Matched Region	
Neur_chan_LBD	59..246	64/242 (26%) 168/242 (69%)	8.3e-71
Neur_chan_memb	253..440	40/291 (14%) 154/291 (53%)	2.6e-52

**Example 48.**

The NOV48 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 48A.

Table 48A. NOV48 Sequence Analysis			
	SEQ ID NO: 373	1118 bp	
NOV48a, CG173066-01 DNA Sequence	GCCCTTAGATCAAGATGCGCTGTAAGTGAAGAGCCCAAGGCGGAGGCTGAGAATCA GAGACATTTTCAGCAGACATCTACAAATCCGAAAGACAAAACATGGTTCAAGCATCCGG GCACAGGCGGTCCACCCGTGGCTCCAAAATGGTCTCCTGGTCCGTGATAGCAAGATC CAGGAAATACTGCAGAGGAAGATGGTGCAGAGATTCTGGCCGAGTTCATGAGCACAT ATGTCATGATGGTATTGGCCCTTGGTTCCGTGGCCCATATGGTTCTAAATAAAAAATA TGGGAGCTACCTTGGTGTCAACTTGGGTTTGGCTTCGGAGTCACCATGGGAGTGCAC GTGGCAGGCCGCATCTCTGGAGCCACATGAACGCAGCTGTGACCTTTGCTAACTGTG CGCTGGGCCGCGTGGCTGGAGGAAGTTCCGGTCTATGTGCTGGGGCAGTTCCTGGG CTCCTTCTGGCGGCTGCCACCATCTACAGTCTCTTCTACACGGCCATTCTCCACTTT TCGGGTGGACAGCTGATGGTGACCGGTCCCGTCCGTACAGCTGGCATTTTTGCCACCT ACCTTCCTGATCACATGACATTGTGGCGGGGCTTCCTGAATGAGGCGTGGCTGACCGG GATGCTCCAGCTGTGCCTCTTCGCCATCACGGACCAGGAGAACACCCAGCACTGCCA GGAACAGAGGCGCTGGTGTATAGGCATCCTCGTGGTCATCATCGGGGTGTCCCTTGGA TGAACACAGGATATGCCATCAACCCGTCCCGGGACCTGCCCCCGCATCTTACCTT CATTGCTGGTTGGGGCAACAGGTCTTCAGTGGCATCATCTACCTGGTCTTCATTGG CTCCACCATCCCACGGGAGCCCTGAAATTGGAGGATTCTGTGGCGTATGAAGACCAC GGGATAACCGTATTGCCCAAGATGGGATCTCATGAACCCACGATCTCTCCCTCACCC CCGTCTCTGTGAGCCCTGCCAACAGATCTTCAGTCCACCCTGCCCCACCCTTACATGA ATCCATAGCCCTAGAGCACTTCTAAGCAGAGATTATTTGTGATCCCATTCCCA ATAAAGCAAGGCTTGT		
	ORF Start: ATG at 100		ORF Stop: TGA at 919
	SEQ ID NO: 374	273 aa	MW at 29820.8kD
NOV48a, CG173066-01 Protein Sequence	MVQASGHRSTRGSKMVSWSVIAKIQEILQRKMVREFLAEFMSTYVMMVFGLSVAHM VLNKKYGSYLGVNLGFGFVMTMGVHVAGRISGAHMNAAVTFANCALGRVPWRKFPVYV LGQFLGSFLAAATYSLFYTAILHFSGGQLMVTGPVATAGIFATYLPDHMTLWRGFLN EAWLTGMLQLCLFAITDQENNPALPGTEALVIGILVVIIGVSLGMNTGYAINPSRDLF PRIFTFIAGWGKQVFRWHHLPLHLHHTGAPEIGGFCGV		

- 5 Further analysis of the NOV48a protein yielded the following properties shown in Table 48B.

Table 48B. Protein Sequence Properties NOV48a	
PSort analysis:	0.8586 probability located in mitochondrial inner membrane; 0.7000 probability located in plasma membrane; 0.6400 probability located in microbody (peroxisome); 0.3568 probability located in mitochondrial intermembrane space

SignalP analysis:	No Known Signal Sequence Predicted
-------------------	------------------------------------

A search of the NOV48a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 48C.

Table 48C. Geneseq Results for NOV48a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV48a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAW87644	A protein with water channel activity - Homo sapiens, 342 aa. [WO9843997-A1, 08-OCT-1998]	1..272 1..268	253/272 (93%) 256/272 (94%)	e-143
AAV70455	Human membrane channel protein-5 (MECHP-5) - Homo sapiens, 341 aa. [WO200012711-A2, 09-MAR-2000]	5..272 3..267	249/269 (92%) 252/269 (93%)	e-140
AAE13275	Human transporters and ion channels (TRICH)-2 - Homo sapiens, 346 aa. [WO200177174-A2, 18-OCT-2001]	1..272 1..272	236/276 (85%) 243/276 (87%)	e-130
ABG27139	Novel human diagnostic protein #27130 - Homo sapiens, 225 aa. [WO200175067-A2, 11-OCT-2001]	49..273 1..225	217/225 (96%) 221/225 (97%)	e-130
ABB57440	Human secreted protein encoding polypeptide SEQ ID NO. 86 - Homo sapiens, 292 aa. [WO200183510-A1, 08-NOV-2001]	29..273 17..258	116/246 (47%) 165/246 (66%)	3e-64

- 5 In a BLAST search of public sequence databases, the NOV48a protein was found to have homology to the proteins shown in the BLASTP data in Table 48D.

Table 48D. Public BLASTP Results for NOV48a				
Protein Accession Number	Protein/Organism/Length	NOV48a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
O14520	Aquaporin 7 (Aquaporin-7 like) (Aquaporin adipose) (AQPap) - Homo sapiens (Human), 342 aa.	1..272 1..268	254/272 (93%) 257/272 (94%)	e-143
BAC05693	Aquaporin adipose - Homo	1..272	253/272 (93%)	e-142

	sapiens (Human), 342 aa.	1..268	256/272 (94%)	
Q8WX69	BA251O17.3 (similar to aquaporin 7) - Homo sapiens (Human), 346 aa.	1..272 1..272	237/276 (85%) 243/276 (87%)	e-130
O54794	Aquaporin 7 - Mus musculus (Mouse), 303 aa.	16..272 1..253	193/257 (75%) 218/257 (84%)	e-108
AAM81581	Aquaporin 7 variant - Rattus norvegicus (Rat), 269 aa.	20..272 4..252	184/253 (72%) 216/253 (84%)	e-106

PFam analysis predicts that the NOV48a protein contains the domains shown in the Table 48E.

Table 48E. Domain Analysis of NOV48a			
Pfam Domain	NOV48a Match Region	Identities/ Similarities for the Matched Region	Expect Value
MIP	27..251	71/247 (29%) 168/247 (68%)	1.5e-56

#### Example 49.

5 The NOV49 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 49A.

Table 49A. NOV49 Sequence Analysis			
	SEQ ID.NO: 375	1461 bp	
NOV49a, CG173085-01 DNA Sequence	GAATTGAAGTGAATGGAACAGAAGCCAAGCAAGGTGGAGTGTGGGTGAGACCCAGAGG AGAACAGTGCCAGGTCACCAGATGGAAACCGAAAAAGAAAGACGGCCAATGTTCCCT GAAACAGCATGTCAGGGTATATCCCTAGTTACCTGGACAAAGACGAGCAGTGTGTC GTGTGTGGGGACAAGGCAACTGGTTATCACTACCGCTGTATCACTTGTGAGGGCTGCA AGGGCTTCTTTTCGCCGACAAATCCAGAAGAACCTCCATCCACCTATTCTGCAAATA TGACAGCTGCTGTGTCATTGACAAGATCACCCGCAATCAGTGCCAGCTGTGCCGCTTC AAGAAGTGCATCGCCGTGGGCATGGCCATGGACTTGGTTCTAGATGACTCGAAGCGGG TGGCCAAGCGTAAGCTGATTGAGCAGAACCGGGAGCGGCGCGGAAGGAGGAGATGAT CCGATCACTGCAGCAGCGACCAGAGCCCACTCCTGAAGAGTGGGATCTGATCCACATT GCCACAGAGGCCCATCGCAGCACAATGCCAGGGCAGCCATTGGAACAGAGGCGGA AATTCCTGCCCGATGACATTGGCCAGTCACCCATTGTCTCCATGCCGACGGAGACAA GGTGGACCTGGAAGCCTTCAGCGAGTTTACCAAGATCATACCCCGGCCATCACCCGT GTGGTGGACTTTGCCAAAAAACTGCCCATGTTCTCCGAGCTGCCTTGCGAAGACCAGA TCATCCTCCTGAAGGGGTGCTGCATGGAGATCATGTCCCTGCGGGCGGCTGTCCGCTA CGACCTGAGAGCGACACCCTGACGCTGAGTGGGGAGATGGCTGTCAAGCGGGAGCAG CTCAAGAATGGCGGCCTGGGCGTAGTCTCCGACGCCATCTTTGAAGTGGGCAAGTCAC TCTCTGCCTTTAACTGGATGACACGGAAGTGGCTCTGCTGCAGGCTGTGCTGCTAAT GTCAACAGACCGCTCGGGCCTGTGTGTGTGGACAAGATCGAGAAGAGTCAGGAGGCG TACCTGCTGGCGTTCGAGCAGCAGCTCAACCACCGCAAACACAACATTCCGCACTTCT GGCCCAAGCTGCTGATGAAGGGTCCGCAGGTCCGGCAGCTTGAGCAGCAGCTTGGTGGA AGCGGGAAGTCTCCAAGGGCCGTTCTTCAGCACCAGAGCCCGAAGAGCCCGCAGCAG CGTCTCCTGGAGCTGCTCCACCGAAGCGGAATTCTCCATGCCCGAGCGGTCTGTGGGG AAGACGACAGCAGTGAGGCGGACTCCCCGAGTCTCTGAGGAGGAACCGGAGGTCTG		

	CGGGGACCTGGCAGGCAATGCAGCCTCTCCCTGAAGCCCCCAGAAGGCCGATGGGGA AGGAGAAGGAGTGCCATACCTTCTCCCAGGCCTCTGCCCCAAGAGCAGGAGGTGCCTG AAAGCTGGGAG		
	ORF Start: ATG at 13		ORF Stop: TGA at 1366
	SEQ ID NO: 376	451 aa	MW at 50612.1kD
NOV49a, CG173085-01 Protein Sequence	MEQKPSKVECGSDPEENSARS PDGNRKRKNGQCSLKTSMGYI PSYLDKDEQCVVCGD KATGYHYRCITCEGCKGFFRRTIQKNLHPTYSKYDSCCVIDK ITRNQCQLCRFKCI AVGMAMDLVLDDSKRVAKRKLIEQNRERRRKEEMIRSLQQRPEPTPEWDLIHIATE HRSTNAQGS HWKQRRKFLPDDIGQSP IVSMPDGD KVDLEAFSEFTKIITPAITRVVDF AKKLPMFSELPCEDQII LLKGCCMEIMSLRAAVRYDPESDTLTLSGEMAVKREQ LKNG GLGVVSDAIFELGKSLSAFNLDDTEVALLQAVLLMSTDRSGLLCVDKIEKSQEAYLLA FEHDVNRKHNI PHFWPKLLMKGPQVRQLEQQQLGEAGSLQGPVLQHQSPKSPQQRLL LLHRSGILHARAVCGEDDSSEADSPSSSEEEPEVCGDLAGNAASP		
	SEQ ID NO: 377		1375 bp
NOV49b, 311531811 DNA Sequence	CACCGGATCCACCATGGAACAGAAGCCAAGCAAGGTGGAGTGTGGGTGAGACCAGAG GAGAACAGTGCCAGGTACCAGATGGAAGCGAAAAAGAAAGAACGGCCAATGTTCCC TGAAAACAGCATGTGAGGGTATATCCCTAGTTACCTGGACAAAGACGAGCAGTGTGT CGTGTGTGGGGACAAGGCAACTGTTATCACTACCGCTGTATCACTTGTGAGGGCTGC AAGGGCTTCTTTCGCGCCACAATCCAGAAGAACCTCCATCCCACCTATTCTTGCAAA ATGACAGCTGCTGTGTCATTGACAAGATCACCCGCAATCAGTGCCAGCTGTGCCGCTT CAAGAAGTGCAATCGCCGTGGGCATGGCCATGGACTTGGTTCTAGATGACTCGAAGCGG GTGGCCAAGCGTAAGCTGATTGAGCAGAACCGGGAGCGGCGGCGGAAGGAGGAGATGA TCCGATCACTGCAGCAGCGACCAAGAGCCACTCCTGAAGAGTGGGATCTGATCCACAT TGCCACAGAGGCCCATCGCAGCACCAATGCCAGGGCAGCCATTGGAAACAGAGGCGG AAATTCCTGCCGATGACATTGGCCAGTCACCCATTGTCTCCATGCCGCGGACGGAGACA AGGTGACCTGGAAGCCTTCAGCGAGTTTACCAAGATCATACCCCGGCCATCACCCG TGTGGTGGACTTTGCCAAAAAACTGCCCATGTTCTCCGAGCTGCCTTGCGAAGACCAG ATCATCTCTGTAAGGGGTGCTGCATGGAGATCATGTCCTGCGGCGGCTGTCCGCT ACGACCCTGAGAGCGACACCTGACGCTGAGTGGGGAGATGGCTGTCAAGCGGGAGCA GCTCAAGAATGGCGGCTGGGCGTAGTCTCCGACGCCATCTTTGAACTGGGCAAGTCA CTCTGCTGCTTTAAGCTGGATGACACGGAAGTGGCTGCTGCTGAGGCTGTGCTGCTAA TGTCACAGACCGCTCGGGCCTGCTGTGTGTGGACAAGATCGAGAAGAGTCAGGAGGC GTACCTGCTGGCGTTTCGAGCACTACGTCAACCACCGCAAACACAACATTCCGCACCTC TGGCCCAAGCTGCTGATGAAGGGTCCGCAAGTCCCGCAGCTTGAGCAGCAGCTTGGTG AAGCGGGAAGTCTCCAAGGGCCGGTCTTTCAGCACCAGAGCCGAAGAGCCCGCAGCA GCGTCTCCTGAGCTGCTCCACCGAAGCGGAATTCTCCATGCCCCGAGCGGCTCTTTGGG GAAGACGACAGCAGTGAGGCGGACTCCCCGAGCTCCTCTGAGGAGGAACCGGAGGTCT GCGAGGACCTGGCAGGCAATGCAGCCTCTCCCGTCGACGGC		
	ORF Start: at 2		ORF Stop: end of sequence
	SEQ ID NO: 378	458 aa	MW at 51408.0kD
NOV49b, 311531811. Protein Sequence	TGSTMEQKPSKVECGSDPEENSARS PDGKRKRKNGQCSLKTSMGYI PSYLDKDEQCV VCGDKATGYHYRCITCEGCKGFFRRTIQKNLHPTYSKYDSCCVIDK ITRNQCQLCRF KKCIAVGMAMDLVLDDSKRVAKRKLIEQNRERRRKEEMIRSLQQRPEPTPEWDLIHI ATEAHRSTNAQGS HWKQRRKFLPDDIGQSP IVSMPDGD KVDLEAFSEFTKIITPAITR VVDFAKKLPMFSELPCEDQII LLKGCCMEIMSLRAAVRYDPESDTLTLSGEMAVKREQ LKNGGLGVVSDAIFELGKSLSAFNLDDTEVALLQAVLLMSTDRSGLLCVDKIEKSQEA YLLAFEHYVNRKHNI PHFWPKLLMKGPQVRQLEQQQLGEAGSLQGPVLQHQSPKSPQQ RLLELLHRSGILHARAVFGEDDSSEADSPSSSEEEPEVCEDLAGNAASPVDG		

- Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 49B.

**Table 49B. Comparison of NOV49a against NOV49b.**

Protein Sequence	NOV49a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV49b	1..451 5..455	447/451 (99%) 447/451 (99%)

Further analysis of the NOV49a protein yielded the following properties shown in Table 49C.

Table 49C. Protein Sequence Properties NOV49a	
PSort analysis:	0.9700 probability located in nucleus; 0.1000 probability located in mitochondrial matrix space; 0.1000 probability located in lysosome (lumen); 0.0000 probability located in endoplasmic reticulum (membrane)
SignalP analysis:	No Known Signal Sequence Predicted

5 A search of the NOV49a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 49D.

Table 49D. Geneseq Results for NOV49a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV49a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAP80926	Sequence of the human thyroid receptor hERBA 8.7 - Homo sapiens, 490 aa. [WO8803168-A, 05-MAY-1988]	1..451 1..490	448/490 (91%) 448/490 (91%)	0.0
AAR26899	HerbA-T sequence - Homo sapiens, 490 aa. [US5144007-A, 01-SEP-1992]	1..451 1..490	446/490 (91%) 447/490 (91%)	0.0
AAY21630	Ligand binding domain of nuclear receptor hTRalpha - Homo sapiens, 410 aa. [WO9926966-A2, 03-JUN-1999]	1..377 1..377	369/377 (97%) 371/377 (97%)	0.0
AAR78318	Human thyroid hormone receptor alpha-1 - Homo sapiens, 410 aa. [US5438126-A, 01-AUG-1995]	1..377 1..377	369/377 (97%) 371/377 (97%)	0.0
AAY21629	Ligand binding domain of nuclear receptor rTRalpha - Rattus sp, 410 aa. [WO9926966-A2, 03-JUN-1999]	1..377 1..377	364/377 (96%) 369/377 (97%)	0.0

In a BLAST search of public sequence databases, the NOV49a protein was found to have homology to the proteins shown in the BLASTP data in Table 49E.

Table 49E. Public BLASTP Results for NOV49a				
Protein Accession Number	Protein/Organism/Length	NOV49a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
AAH35137	Similar to thyroid hormone receptor - Homo sapiens (Human), 451 aa.	1..451 1..451	448/451 (99%) 448/451 (99%)	0.0
P10827	Thyroid hormone receptor alpha (C-erbA-alpha) (c-erbA-1) (EAR-7) (EAR7) - Homo sapiens (Human), 490 aa.	1..451 1..490	448/490 (91%) 448/490 (91%)	0.0
O97716	Thyroid hormone receptor alpha (C-erbA-alpha) (c-erbA-1) - Sus scrofa (Pig), 506 aa.	1..445 1..484	434/484 (89%) 439/484 (90%)	0.0
I57696	c-erbA-alpha-2-related protein - rat, 492 aa.	1..451 1..492	435/492 (88%) 441/492 (89%)	0.0
S14418	thyroid hormone receptor alpha-3 - mouse, 413 aa (fragment).	1..413 1..413	407/413 (98%) 410/413 (98%)	0.0

PFam analysis predicts that the NOV49a protein contains the domains shown in the Table 49F.

Table 49F. Domain Analysis of NOV49a			
Pfam Domain	NOV49a Match Region	Identities/ Similarities for the Matched Region	Expect Value
zf-C4	51..128	50/78 (64%) 71/78 (91%)	2e-52
hormone_rec	223..408	58/212 (27%) 136/212 (64%)	7.2e-34

## 5 Example 50.

The NOV50 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 50A.

Table 50A. NOV50 Sequence Analysis			
	SEQ ID NO: 379	2174 bp	
NOV50a,	GCCCTTTATCGCCGGAATTCATGTGCAATACCAACATGTCTGTACCTACTGATGGTGC		

CG173095-01 DNA Sequence	<p>TGTAACCACCTCACAGATTCCAGCTTCGGAACAAGAGACCCTGGTTAGACCAAAGCCA  TTGCTTTTGAAGTTATTAAAGTCTGTTGGTGCACAAAAGACACTTATACTATGAAAG  AGAGATGGAGTTTCACTATGTTGCCAGGCTGGTCTGGAACCTCTGGGCTCAAGGGAT  CTGCTTACCTCGGCCTCCTAAAGTGCTAGATTACAGGTTCTTTTTATCTTGGCCAG  TATATTATGACTAAACGATTATATGATGAGAAGCAACAACATATTGTATATTGTTCAA  ATGATCTTCTAGGAGATTGTTTGGCGTGCCAAGCTTCTCTGTGAAAGAGCACAGGAA  AATATATACCATGATCTACAGGAAGCTTGGTAGTAGTCAATCAGCAGGAATCATCGGAC  TCAGGTACATCTGTGAGTGAGAACAGGTGTCACCTTGAAGGTGGGAGTGATCAAAAGG  ACCTGTGACAAGAGCTTCAGGAAGAGAAACCTTCATCTTCACATTGGTTTCTAGACC  ATCTACCTCATCTAGAAAGAGAGCAATTAGTGAGACAGAAGAAAATTCAGATGAATTA  TCTGGTGAACGACAAAGAAAACGCCACAAATCTGATAGTATTTCCCTTTCCCTTTGATG  AAAGCCTGGCTCTGTGTGTAATAAGGGAGATATGTTGTGAAAGAAGCAGTAGCAGTGA  ATCTACAGGGACGCCATCGAATCCGGATCTTGATGCTGGTGTAAAGTGAACATTCAGGT  GATTGGTTGGATCAGGATTCAGTTTCAGATCAGTTTAGTGTAAGATTTGAAGTTGAAT  CTCTCGACTCAGAAGATTATAGCCTTAGTGGAAGAAGGACAAGAAGCTTCAGATGAAGA  TGATGAGGTATATCAAGTTACTGTGTATCAGGCAGGGGAGAGTGATACAGATTCATTT  GAAGAAGATCCTGAAATTTCTTAGCTGACTATTGGAAATGCATTCATGCAATGAAA  TGAATCCCCCCTTCCATCACATTGCAACAGATGTTGGGCCCTTCGTGAGAATTGGCT  TCCTGAAGATAAAGGGAAAGATAAAGGGGAAATCTCTGAGAAAGCCAACTGGAAGAAC  TCAACACAAGCTGAAGAGGGCTTTGATGTTCTGATTTGTAAGAACTATAGTGAATG  ATTCCAGAGAGTCATGTGTTGAGGAAAATGATGATAAAATTACACAAGCTTCACAATC  ACAAGAAAGTGAAGACTATTCTCAGCCATCAACTTCTAGTAGCATTTATTTATAGCAGC  CAAGAAGATGTGAAAGAGTTTGAAGGGGAAGAAACCCAAGACAAGAAGAGAGTGTGG  AATCTAGTTTGGCCCTTAATGCCATTGAACCTTGTGTGATTTGTCAAGGTCGACCTAA  AAATGGTTGCATTGTCCATGGCAAAACAGGACATCTTATGGCCTGCTTTACATGTGCA  AAGAAGCTAAAGAAAAGGAATAAGCCCTGCCAGTATGTAGACAACCAATTCAAATGA  TTGTGCTAACTTATTTCCCTAGTTGACCTGTCTATAAGAGAATTATATATTTCTAAC  TATATAACCTTAGGAATTTAGACAACCTGAAATTTATTCACATATATCAAAGTGAGAA  AATGCCTCAATTCACATAGATTTCTTCTTTAGTATAATTGACCTACTTTGGTAGTG  GAATAGTGAATACTTACTATAATTTGACTTGAATATGTAGCTCATCTTTACACCAAC  TCCTAATTTTAAATAATTTCTACTCTGTCTTAAATGAGAAGTACTTTGGTTTTTTTTT  CTTAAATATGTATATGACATTTAAATGTAACCTATTATTTTTTTTGAGACCGAGTCTT  GCTCTGTTACCCAGGCTGGAGTGCAGTGGGTGATCTTGGCTCACTGCAAGCTCTGCCC  TCCCCGGGTTCCGACCATTCTCCTGCCTCAGCCTCCCAATTAGCTTGGCCTACAGTCA  TCTGCCACCACACTGGCTAATTTTTTTGACTTTTAGTAGAGACAGGTTTCCACGTG  TTAGCCAGGATGGTCTCGATCTCCTGACCTCGTGATCCGCCACCTCGGCCTCCCAAA  GTGCTGGGATTACAGGCATGAGCCACCG</p>
	<div>ORF Start: ATG at 21</div> <div>ORF Stop: TAG at 1587</div>
	<div>SEQ ID NO: 380</div> <div>522 aa</div> <div>MW at 58895.6kD</div>
NOV50a, CG173095-01 Protein Sequence	<p>MCNTNMSVPTDGAVTTSQIPASEQETLVRPKPLLLKLLKSVGAQKDYTMKERSFTM  LPRLVWNSWAQIGICLPRPPKVLDDLQVLFYLGQYIMTKRLYDEKQQHIVYCSNDLLGDL  FGVPFSFSVKEHRKIYTMIRNLVVVNQQESSDSGTSVSENRCHEGGSDQKDLVQELQ  EEKPSSSHLVSRPSTSSRKRAISETENSDELSEGERQKRHKSDSISLSFDESALCV  IREICERSSSSESTGTPSNPDLDAGVSEHSGDWLDQDSVSDQFSVEFEVESLDSY  SLSEEGQELSDDEDEVYQVTVYQAGESDSDSFEEDPEISLADYWKCTSCNEMNPPLPS  HCNRCWALRENWLPEDKGDKGEISEKAKLENSTQAEFGFDPDCKKTIVNDSRESCV  EENDDKITQASQSQESSEDYQPSSTSSSIYSSQEDVKEFEREETQDKESVESLPLN  AIEPCVICQGRPKNGCIVHGKTGHLMACTCAKLLKRNKPCPVCRQPIQMIVLTYFP</p>
	<div>SEQ ID NO: 381</div> <div>1607 bp</div>
NOV50b, CG173095-02 DNA Sequence	<p>GCCCTTTATCGCCGAATTCATGTGCAATACCAACATGCTGTACCTACTGATGGTGC  TGTAACCACCTCACAGATTCCAGCTTCGGAACAAGAGACCCTGGTTAGACCAAAGCCA  TTGCTTTTGAAGTTATTAAAGTCTGTTGGTGCACAAAAGACACTTATACTATGAAAG  AGAGATGGAGTTTCACTATGTTGCCAGGCTGGTCTGGAACCTCTGGGCTCAAGGGAT  CTGCTTACCTCGGCCTCCTAAAGTGCTAGATTACAGGTTCTTTTTATCTTGGCCAG  TATATTATGACTAAACGATTATATGATGAGAAGCAACAACATATTGTATATTGTTCAA  ATGATCTTCTAGGAGATTGTTTGGCGTGCCAAGCTTCTCTGTGAAAGAGCACAGGAA  AATATATACCATGATCTACAGGAAGCTTGGTAGTAGTCAATCAGCAGGAATCATCGGAC  TCAGGTACATCTGTGAGTGAGAACAGGTGTCACCTTGAAGGTGGGAGTGATCAAAAGG</p>

	ACCTTGTACAAGAGCTTCAGGAAGAGAAACCTTCATCTTCACATTTGGTTTCTAGACC ATCTACCTCATCTAGAAGGAGAGCAATTAGTGAGACAGAAGAAAATTCAGATGAATTA TCTGGTGAACGACAAAGAAAACGCCACAAATCTGATAGTATTTCCCTTTCCTTTGATG AAAGCTTGGCTCTGTGTGTAATAAGGGAGATATGTTGTGAAAGAAGCGGTAGCAGTGA ATCTACAGGGACGCCATCGAATCCGGATCTTGATGCTGGTGTAAAGTGAACATTCAGGT GATTGGTTGGATCAGGATTCAGTTTCAGATCAGTTTAGTGTAGAATTTGAAGTTGAAI CTCTCGACTCAGAAGATTATAGCCTTAGTGAAGAAGACAAGAACTCTCAGATGAAGA TGATGAGGTATATCAAGTTACTGTGTATCAGGCAGGGGAGAGTGATACAGATTCATTT GAAGAAGATCCTGAAATTTCTCAGCTGACTATTGGAAATGCACTTCATGCAATGAAA TGAATCCCCCTTCCATCAGATTGCAACAGATGTTGGGCCCTTCGTGAGAATTGGCT TCCTGAAGATAAAGGGAAAGATAAAGGGGAAATCTCTGAGAAAGCCAAACTGGAAGAAC TCAACACAAGCTGAAGAGGGCTTTGATGTTCTTGATTGTAAGAAAATATAGTGAATG ATTCCAGAGAGTCATGTGTTGAGGAAAATGATGATAAAATACACAAGCTTCACAATC ACAAGAAAGTGAAGACTATTCTCAGCCATCAACTTCTAGTAGCATTATTTATAGCAGC CAAGAAGATGTGAAAGAGTTTGAAAGGGAAGAAACCAAGACAAGAAAGAGAGTGTGG AATCTAGTTTGGCCCTTAATGCCATTGAACCTTGTGTGATTTGCCAAGGTCGACCTAA AAATGGTTGCATTGTCCATGGCAAAACAGGACATCTTATGGCCTGCTTTACATGTGCA AAGAAGCTAAAGAAAAGGAATAAGCCCTGCCCTGTATGTAGACAACCAATTCAAATGA TTGTGCTAACTATTTCTCCTGACTGCAGCCAAGCTAATTC		
	ORF Start: ATG at 21		ORF Stop: TGA at 1587
	SEQ ID NO: 382	522 aa	MW at 58857.5kD
NOV50b, CG173095-02 Protein Sequence	MCNTNMSVPTDGAVTTSQIPASEQETLVRPKPLLLKLLKSVGAQKDYTMKERWSFTM LPRLVWNWAQIGICLPRPPKVLDDLQVLFYLGQYIMTKRLYDEKQQHIIVYCSNDLLGDL FGVPSFSVKEHRKIYTMIRNLVVVNQQESSDSGTSVSENRCHEGGSDQKDLVQELQ EEKPSSSHLVSRPSTSSRRRAISETEENSDELSEGERQKRKHSISLSFDESLALCV IREICCERSGSSESTGTPSPNDLDAGVSEHSGDWLDQDSVSDQFSVEFEVESLDSY SLSEEGQELSDDEDDEVYQVTYVYQAGESDTSFEEDPEISSADYWKCTSCNEMNPPLPS HCNRCWALRENWLPEDKGKDKGEISEKAKLENSTQABEGFDVPDCKKTIIVNDSRESCV EENDDKITQASQSQESSEDYSQPSTSSSIYSSQEDVKEFEREETQDKEESVSSLPLN AIEPCVICQGRPKNGCIVHGKTGHLMACFTCAKLLKRNKPCPVCROPIQMIVLTYFS		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 50B.

Table 50B. Comparison of NOV50a against NOV50b.		
Protein Sequence	NOV50a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV50b	1..521	518/521 (99%)
	1..521	519/521 (99%)

5 Further analysis of the NOV50a protein yielded the following properties shown in Table 50C.

Table 50C. Protein Sequence Properties NOV50a	
PSort analysis:	0.6000 probability located in nucleus; 0.3000 probability located in microbody (peroxisome); 0.1000 probability located in mitochondrial matrix space; 0.1000 probability located in lysosome (lumen)
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV50a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 50D.

<b>Table 50D. Geneseq Results for NOV50a</b>				
<b>Geneseq Identifier</b>	<b>Protein/Organism/Length [Patent #, Date]</b>	<b>NOV50a Residues/ Match Residues</b>	<b>Identities/ Similarities for the Matched Region</b>	<b>Expect Value</b>
AAO15376	Human Dm2 (Hdm2) protein - Homo sapiens, 491 aa. [US2002045192-A1, 18-APR-2002]	1..522 1..491	490/522 (93%) 491/522 (93%)	0.0
AAE22654	Human Ring finger E3 ubiquitin ligase (Mdm2) protein - Homo sapiens, 491 aa. [WO200197830-A1, 27-DEC-2001]	1..522 1..491	490/522 (93%) 491/522 (93%)	0.0
AAB48284	Human MDM2 protein - Homo sapiens, 491 aa. [WO200075184-A1, 14-DEC-2000]	1..522 1..491	490/522 (93%) 491/522 (93%)	0.0
AAY96567	MDM2 oncoprotein - Homo sapiens, 491 aa. [WO200031238-A2, 02-JUN-2000]	1..522 1..491	490/522 (93%) 491/522 (93%)	0.0
AAW94304	Human MDM2 - Homo sapiens, 491 aa. [US5858976-A, 12-JAN-1999]	1..522 1..491	490/522 (93%) 491/522 (93%)	0.0

- 5 In a BLAST search of public sequence databases, the NOV50a protein was found to have homology to the proteins shown in the BLASTP data in Table 50E.

<b>Table 50E. Public BLASTP Results for NOV50a</b>				
<b>Protein Accession Number</b>	<b>Protein/Organism/Length</b>	<b>NOV50a Residues/ Match Residues</b>	<b>Identities/ Similarities for the Matched Portion</b>	<b>Expect Value</b>
Q00987	Ubiquitin-protein ligase E3 Mdm2 (EC 6.3.2.-) (p53-binding protein Mdm2) (Oncoprotein Mdm2) (Double minute 2 protein) (Hdm2) - Homo sapiens (Human), 491 aa.	1..522 1..491	490/522 (93%) 491/522 (93%)	0.0
P56951	Ubiquitin-protein ligase E3 Mdm2 (EC 6.3.2.-) (p53-binding protein Mdm2) (Oncoprotein Mdm2) (Double minute 2 protein) (Edm2) -	1..522 1..491	463/522 (88%) 479/522 (91%)	0.0

	Equus caballus (Horse), 491 aa.			
Q9GMZ6	MDM2 - Canis familiaris (Dog), 487 aa.	1..522 1..487	456/522 (87%) 466/522 (88%)	0.0
P56950	Ubiquitin-protein ligase E3 Mdm2 (EC 6.3.2.-) (p53-binding protein Mdm2) (Oncoprotein Mdm2) (Double minute 2 protein) (Cdm2) - Canis familiaris (Dog), 487 aa.	1..522 1..487	454/522 (86%) 464/522 (87%)	0.0
Q95KN5	MDM2 - Canis familiaris (Dog), 487 aa.	1..522 1..487	453/522 (86%) 463/522 (87%)	0.0

PFam analysis predicts that the NOV50a protein contains the domains shown in the Table 50F.

Table 50F. Domain Analysis of NOV50a			
Pfam Domain	NOV50a Match Region	Identities/ Similarities for the Matched Region	Expect Value
MDM2	30..126	56/97 (58%) 82/97 (85%)	1e-39
zf-RanBP	330..359	9/32 (28%) 26/32 (81%)	3.6e-08
zf-C3HC4	469..509	14/55 (25%) 31/55 (56%)	0.81

### Example 51.

5 The NOV51 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 51A.

Table 51A. NOV51 Sequence Analysis			
	SEQ ID NO: 383	2066 bp	
NOV51a, CG173173-01 DNA Sequence	TATTTCAGAAAGCTTCAAGAACAAGCTGGAGAAGGGAAGAGTTATTCCTCCATATTCA CCTGCTTCAACTACTATTCTTATTGGGAATGGACAATGGAATGTTCTCTGGTTTTATC ATGATCAAAAACCTCCTTCTCTTTGTATTTCCATGAACCTATCCAGTCACTTTGGCT TTTCACAGGTGCCAACCAGTTCAGTGAAAGATGAGACCAATGACAACATCACGATATT TACCAGGATCTTGGATGGGCTCTTGGATGGCTACGACAACAGACTTCGGCCCCGGGCTG GGAGAGCGCATCACTCAGGTGAGGACCGACATCTACGTCACCAGCTTCGGCCCCGGTGT CCGACACGGAAATGGAGTACACCATAGACGTGTTTTCCGACAAGGCTGGAAGATGA AAGGCTTCGGTTTAAGGGGCCCATGCAGCGCCTCCCTCTCAACACGTTCTTCCACAAC GGGAAGAAGTCCATCGCTCACAAACATGACCACGCCCCAACAAGCTGCTGCGGCTGGAGG ACGACGGCACCCCTGCTCTACACCATGCGCTTGACCATCTCTGCAGAGTGCCCCATGCA GCTTGAGGACTTCCCGATGGATGCGCACGCTTGCCCTTGAAATTTGGCAGCTATGCG TACCCTAATTCTGAAGTCGTTTACGTCCTGGACCAACGGCTCCACCAAGTCGGTGGTGG TGGCGGAAGATGGCTCCAGACTGAACCACTGATGGGGCAGACGGTGGGCAC TGAGAACATCAGCACCGACAGGCGAATACACAATCATGACAGCTCACTCCACCTG AAAAGGAAGATTGGCTACTTTGTCATCCAGACCTACCTTCCCTGCATAATGACCGTGA		

	TCTTATCACAGGTGTCCTTTTGGCTGAACCGGGAATCAGTCCCAGCCAGGACAGTCTTT TGGGGTCACCACGGTGCTGACCATGACGACCCTCAGCATCAGCGCCAGGAACCTCTCTG CCCAAAGTGGCCTACGCCACCGCCATGGACTGGTTTCATAGCTGTGTGCTATGCCCTCG TCTTCTCGGCGCTGATAGAGTTTGCCACGGTCAATTACTTTACCAAGAGAGGGCTGGGC CTGGGATGGCAAAAAGCCTTGGAAGCAGCCAAGATCAGAAAAAGCGTGAAGTCATA CTAAATAAGTCAACAAACGCTTTTACAACCTGGGAAGATGTCTCACCCCCAACATTC CGAAGGAACAGACCCCAGCAGGGACGTGCAATACAACCTCAGTCTCAGTAAAACCTC TGAAGAGAAGACTTCTGAAAGCAAAAAGACTTACAACAGTATCAGCAAAATTGACAAA ATGTCCCAGAATCGTATTTCCAGTCTTGTTTCGGCACTTTCAACTTAGTTTACTGGGCAA CGTATTTGAATAGGGAGCCGGTGATAAAGGAGCCGCCTCTCCAAAATAACCGGCCAC ACTCCCAAACTCCAAGACAGCCATACTTCCAGCGAAATGGTACCAAGGAGAGGTTTGTG CTCACAGGGACTCTCCATATGTGAGCACTATCTTTTCAGGAAATTTTTCATGTTTAAAT AATATGTACAAATAATATTGCCTTGATGTTTCTATATGTAACCTCAGATGTTTCCAAG ATGTCCCATTGATAATTCGAGCAAAACAACCTTCTGGAAAAACAGGATACGATGACTGA CACTCAGATGCCCAGTATCATACGTTGATAGTTTACAACAAGATACGTATATTTTAA ACTGCTTCAAGTGTTTACCTAACAAATGTTTATACTTCAAATGTCAATTTTCATACAAA TTTTCCAGTGAATAAATATTTTAGGAAACTCTCCATGATTATTAGAAGACCAACTAT ATTGCGAGAAACAGAGATCATAAAGAGCACGTTTCCATTATGAGGAACTTGGACAT TTATGTACAAATGAATTGCCTTTGATAATCTTACTGTTCTGAAATTAGGAAAGTAC TTGCATGATCTTACACGAAGAAATAGAATAGGCAAACTTTTATGTAGGCAGATTAATA ACAGAAATACATCATATGTTAGATACACAAAATATT		
	ORF Start: ATG at 87		ORF Stop: TAA at 1440
	SEQ ID NO: 384	451 aa	MW at 50844.0kD
NOV51a, CG173173-01 Protein Sequence	MDNGMFSGFMIKNLLFCISMNLSHFGFSQVPTSSVKDETNDNITIFTRILDGLLD GYDNRLRPLGERITQVRTDIYVTSFGPVSDTEMEYIDVFFRQGWKDERLRFKGP RLPLNTFFHNGKKSIAHNMTPNKLRLLEDDGTLTYMRLTISAECPMQLEDPFMDAH ACPLKFGSYAYPNSEVVYVWNGSTKSVVVAEDGSRLNQYHLMGQTVGTENISTSTGE YTIMTAHFHLKRKIGYFVIQTYLPCIMTVILSQVSFWLNRESVPARTVFGVTTVLMT TSLISARNSLPKVAYATAMDWFIACVAFVFSALIEFATVNYFTKRGAWDGGKALEA AKIKKKREVILNKSTNAFTTGKMSHPPNIPKEQTPAGTSNNTSVSVKPSSEKTSSESK TYSISKIDKMSRIVFPVLFGTFNLVYWATYLNREPVIKGAASPK		

Further analysis of the NOV51a protein yielded the following properties shown in Table 51B.

Table 51B. Protein Sequence Properties NOV51a	
PSort analysis:	0.7073 probability located in microbody (peroxisome); 0.7000 probability located in plasma membrane; 0.4477 probability located in mitochondrial inner membrane; 0.2000 probability located in endoplasmic reticulum (membrane)
SignalP analysis:	Cleavage site between residues 32 and 33

- 5 A search of the NOV51a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 51C.

Table 51C. Geneseq Results for NOV51a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV51a Residues/ Match	Identities/ Similarities for the Matched	Expect Value

		<b>Residues</b>	<b>Region</b>	
AAR59864	Human GABA receptor alpha5 subunit - Homo sapiens, 462 aa. [WO9413799-A, 23-JUN-1994]	1..451 1..462	449/462 (97%) 450/462 (97%)	0.0
AAR31186	GABA-A receptor alpha-5 subunit - Homo sapiens, 462 aa. [WO9222652-A, 23-DEC-1992]	1..451 1..462	449/462 (97%) 450/462 (97%)	0.0
AAR59862	Human GABA receptor alpha2 subunit - Homo sapiens, 451 aa. [WO9413799-A, 23-JUN-1994]	39..444 32..447	312/419 (74%) 347/419 (82%)	0.0
AAR31184	GABA-A receptor alpha-2 subunit - Homo sapiens, 451 aa. [WO9222652-A, 23-DEC-1992]	39..444 32..447	312/419 (74%) 347/419 (82%)	0.0
ABG26224	Novel human diagnostic protein #26215 - Homo sapiens, 547 aa. [WO200175067-A2, 11-OCT-2001]	29..446 102..542	310/441 (70%) 345/441 (77%)	e-177

In a BLAST search of public sequence databases, the NOV51a protein was found to have homology to the proteins shown in the BLASTP data in Table 51D.

<b>Table 51D. Public BLASTP Results for NOV51a</b>				
<b>Protein Accession Number</b>	<b>Protein/Organism/Length</b>	<b>NOV51a Residues/ Match Residues</b>	<b>Identities/ Similarities for the Matched Portion</b>	<b>Expect Value</b>
P31644	Gamma-aminobutyric-acid receptor alpha-5 subunit precursor (GABA(A) receptor) - Homo sapiens (Human), 462 aa.	1..451 1..462	449/462 (97%) 450/462 (97%)	0.0
B34130	gamma-aminobutyric acid/benzodiazepine receptor alpha-5 chain precursor - rat, 464 aa.	1..451 1..464	427/464 (92%) 437/464 (94%)	0.0
P19969	Gamma-aminobutyric-acid receptor alpha-5 subunit precursor (GABA(A) receptor) - Rattus norvegicus (Rat), 464 aa.	1..451 1..464	427/464 (92%) 437/464 (94%)	0.0
P26048	Gamma-aminobutyric-acid receptor alpha-2 subunit precursor (GABA(A) receptor) - Mus musculus (Mouse), 451 aa.	39..444 32..447	313/419 (74%) 348/419 (82%)	0.0
P23576	Gamma-aminobutyric-acid receptor alpha-2 subunit precursor	39..444 32..447	313/419 (74%) 347/419 (82%)	e-180

	(GABA(A) receptor) - Rattus norvegicus (Rat), 451 aa.			
--	---	--	--	--

Pfam analysis predicts that the NOV51a protein contains the domains shown in the Table 51E.

Table 51E. Domain Analysis of NOV51a			
Pfam Domain	NOV51a Match Region	Identities/ Similarities for the Matched Region	Expect Value
Neur_chan_LBD	49..246	68/267 (25%) 163/267 (61%)	9e-60
Neur_chan_memb	253..434	39/291 (13%) 162/291 (56%)	1.6e-58

### Example 52.

The NOV52 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 52A.

Table 52A. NOV52 Sequence Analysis			
	SEQ ID NO: 385	2266 bp	
NOV52a, CG51213-01 DNA Sequence	CTCGGGCCTGGGGCTCTGCCTGAACAACCGGCCCCAGACAGGACTTTGTGTACCCG ACAGTGGCACCGGGCCAAGCCTACGATGCAGATGAGCAATGCCGCTTTCAGCATGGAG TCAAATCGCGTCAGTTGGTGCTACAAACGGGTCTGTGTCCCTTTGGGTTCGCGCCAG AGGGTGTGGACGGAGCCTGGGGGCCGTGGACTCCATGGGGCGACTGCAGCCGGACCTG TGGCGGCGGCGTGTCTCTTCTAGCCGTCAGTGCAGACAGCCCAAGGCCAACCATCGGG GGCAAGTACTGTCTGGGTGAGAGAAGGCGGCACCGCTCCTGCAACACGGATGACTGTC CCCCTGGCTCCCAGGACTTCAGAGAAGTGCAGTGTCTGAATTTGACAGCATCCCTTT CCGTGGGAAATTCTACAAGTGAAAAACGTACCGGGAAGGTGAGTGTGGGACTCCAAA GGCTGTGGGGCCGTGAAGGGCAGCCGTGGGAGTGTCCAGCAGCAGGTGGATGAATGCA GCATCCCGGGGTCTGCCATGAGCCCTGTCCCACCGGGGAGACAGAGTACCTGGGAT ACGGTACCATGGGGGTCAACGTGACGCTGGGAGCCCCACTCCCTCTGCCCAAGCTG CCCTTCCTCTTGGGTCTGGGGTCTGTCCCTCTTGGCCTCACTCCCCAGGGAGCAAGC AAAGAGTTCCGGGTGGCCTGGCCGTGGTGTGACGGGGCCGTGCCCCCAGGGGGCG TGAAGGCCTGCTCGCTCAGTGCCTAGCGGAAGGCTTCAACTTCTACACGGAGAGGGC GGCAGCCGTGGTGGACGGGACACCCTGCCGTCCAGACACGGTGGACATTTGCGTCAGT GGCGAATGCAAGCACGTGGGCTGCGACCGAGTCTGGGCTCCGACCTGCGGGAGGACA AGTGCCGAGTGTGTGGCGGTGACGGCAGTGCC'TGCGAGACCATCGAGGGCGTCTTCAG CCCAGCCTCACCTGGGGCCGGGTACGAGGATGTGCTGTGGATTCCCAAAGGCTCCGTC CACATCTTCATCCAGGATCTGAACCTCTCTCTCAGTCACTTGCCCTGAAGGGAGACC AGGAGTCCCTGCTGCTGGAGGGGCTGCCTGGGACCCCCAGCCCCACCGTCTGCCTCT AGCTGGGACCACCTTTCAACTGCGACAGGGGCCAGACCAGGTCCAGAGCCTCGAAGCC CTGGGACCGATTAATGCATCTCTCATCGTCATGGTGTCTGGCCCGGACCGAGCTGCCTG CCCTCCGCTACCGCTTCAATGCCCCATCGCCCGTGAATGCTGCCCCCTACTCCTG GCACTATGCGCCCTGGACCAAGTGCTCGGCCAGTGTGCAGGCGGTAGCCAGGTGCAG GCGGTGGAGTGCCGCAACCAGCTGGACAGCTCCGCGGTGCGCCCCCTACTGTCAGTG CCCACAGCAAGCTGCCCCAAAGGCAGCGCGCTGCAACACGGAGCCTTGCCCTCCAGA CTGGGTGTAGGGAAGTGGTCTGCTGTCAGCCGACGTGCGATGCAGGCGTGCAGCAGC CGCTCGGTGCTGTGCCAGCGCCGCTCTCTGCCCGGAGGAGAAGGCGCTGGACGACA GCGCATGCCCCGAGCCGCGCCACCTGTACTGGAGGCCTGCCACGGCCCCACTTGCCC		

	TCCGGAGTGGGCGGCCCTCGACTGGTCTGAGTGCACCCAGCTGCGGGCCGGGCCTCGCCACCGCGTGGTCTTTTGCAAGAGCGCAGACCACCGGCCACGCTGCCCCGGCGCCTGCTCACCCGCGCCAAAGCCACCGGCCACCATGCGCTGCAACTTGCGCCGCTGCCCCCGGCCGCTGGGTGGCTGGCGAGTGGGGTGAGTGCTCTGCACAGTGCGGGCGTCTGGGCAGCGGCAGCGCTCGGTGCGCTGCACCAGCCACACGGGCCAGGCGTCGCACGAGTGACCGAGGCCCTGCGGCCGCCACCACGCAGCAGTGTGAGGCCAAGTGCACAGCCCAACCCCCGGGACGGCCCTGAAGAGTGCAAGGATGTGAACAAGGTCGCCTACTGCCCCCTGTGCTCAAATTTAGTTCTGTCAGCCGAGCCTACTTCCGCCAGATGTGCTGCAAAACCTGCCAGGGCCACTAGGGGGCGCGCGGCCACCCGGAGCCACAGCTGGCGGGGTCTCCGCCCGCAGCCCTGCAGCTGGGCCGGCCAGAGGGGGCCCCGGGGGGCGGGAAGTGGGAGGGAAGGG		
	ORF Start: ATG at 589		ORF Stop: TAG at 2158
	SEQ ID NO: 386	523 aa	MW at 56126.2kD
NOV52a, CG51213-01 Protein Sequence	MGVQRDAGSPHSLCPSPSSWVWGLSLLASLPQGASKEFRGGLARGVTGP PPGGVKACSLTCLAEGFNFYTERAAVVDGTPCRPDTVDICVSGECKHVGCDRLVGLSRLREDKCRVCGGDGSACETIEGVFSPASPGAGYEDVWVWIPKGSVHIFIQDLNLSLSHLALKGQDESLLEGLPGTTPQPHRLPLAGTTFQLRQGPDPQVQSLEALGPINASLIVMVLARTELPALRYRFNAPIARDSLPPYSWHYAPWTKCSAQCAAGSSQVQAVECRNQLDSSAVAPHYCSAHSKLPKRQRACNTEPCPPDWVGNWSLCSRSCDAGVRSRVSVCQRRVSAAEKALDDSAQPQPRPPVLEACHGPTCPPEWAALDWSECTPSCGPGLRHRVVLCKSADHRATLPPAHCS PAAKPPATMRCNLRRCPARWVAGEWGECSAQCGVGQRQSRVCTSHTGQASHCTEALRPPTTQQCEAKCDSPTPGDGPPECKDVNVKVA YCPLVLKFQFCSTRAYFRQMCKKTCQH		
	SEQ ID NO: 387	1866 bp	
NOV52b, CG51213-07 DNA Sequence	TCCATAAATGGAGCTTATTGGGAGAGTATAAGTCACAGGCCATGCCCGCAAGGGGATGCACGAAGACCCACCGCGAGCCAGGAAGGGAGCACCGGGCTCTCTGCTCTGGGACCGGCAGTGAGCCGGACATCTGGGTCTCCCAAGCCGGCGGGCTGCCCCAGGAGGAAGGGAGGGGGGCGAGCCTGAGCGGGCACCTCGGCCCGCAGGAGGTCTGCAGCGAGCTGTGGTGTCTGAGCAAGAGCAACCGGTGCATCACCACAGCATCCCGGCCCGCAGGGCACGCTGTGCCAGACGCACACCATCGACAAGGGGTGGTGTCTACAAACGGGTCTGTGTCCCTTTGGGTGCGGCCAGAGGGTGTGGACGGAGCCTGGGGGCCGTGGACTCCATGGGGCGACTGCAGCCGACCTGTGGCGGCGGCGTGTCTCTTTAGCCGTCACTGCGACAGCCCCAGGCCAACCATCGGGGGCAAGTACTGTCTGGGTGAGAGAAGGCGGCACCGCTCCTGCAACACGGATGACTGTCCCCCTGGCTCCAGGACTTCAGAGAAGTGCAGTGTCTGAATTTGACAGCATCCCTTTCCGTGGGAAATCTACAAGTGGAAAACGTACCGGGAGGGGGCGGTGAAGGCCTGCTCGCTCAGTGCCTAGCGGAAGGCTTCAACTTCTACACGGAGAGGGCGGCAGCCGTGGTGGACGGGACACCTGCCGTCCAGACACGGTGGACATTTGCGTCAGTGGCAATGCAAGCAGTGGGTGCGACCGAGTCTGGGTCCGACCTGCGGGAGGACAA GTGCCGAGTGTGTGGCGGTGACGGCAGTGCCTGCGAGACCATCGAGGGCGTCTTCAGCCAGCCTCACCTGGGGCCGGGTACGAGGATGTGCTCTGGATTCCCAAAGGCTCCGTCCACATCTTCATCCAGGATCTGAACCTCTCTCTCAGTCACTTGCCCTGAAGGGAGACCA GGAGTCCCTGCTGCTGGAGGGGCTGCCCGGGACCCCCAGCCCCACCGTCTGCCTCTAGCTGGGACCACCTTTCAACTGCGACAGGGGCCAGACCAGGTCCAGAGCCTCGAAGCCCTGGGACCGATTAATGCATCTCTCATCGTCATGGTGCTGGCCCGGACCGAGCTGCCTGCTCCTCCGCTACCGCTTCAATGCCCCCATCGCCGTGACTCGCTGCCCCCTACTCCTGGCACTATGCGCCCTGGACCAAGTGCTCGGCCAGTGTGCAGGCGGTAGCCAGGTGCAGGCGGTGGAGTGCCGCAACAGCTGGACAGCTCCGCGGTGCCCCCACTACTGCAGTGC CCACAGCAAGCTGCCAAAAGGCAGCGCGCTGCAACACGGAGCCTTGCCCTCCAGACTGGGTGTAGGGAAGTGGTCTGCTGCTGCGAGCCGAGCTGCGATGCAGGCGTGCGCAGCGCTCGGTGCTGCGCAGCGCGCTCTCTGCGCGGAGGAGAAGGCGCTGGACGACAGCGCATGCCCGCAGCCGCGCCACCTGTACTGGAGGCCTGCCACGGCCCCACTTGCCCTCCGGAGTGGGCGGCCCTCGACTGGTCTGAGTGACCCCCAGCTGCGGGCGGGCCTCCGCCACCGGTGGTCTTTGCAAGAGCGCAGACCACCGCGCCACGCTGCCCGCGGCACTGCTCACCCGCGGCCAAGCCACCGGCCACCATGCGCTGCAACTTGCGCGCTGCCCTCCCGGCCCGTGGTGGTGGCGAGTGGGGTGAGTGTCTGCAAGTGGCGGTGCGGCTCGGGAGCGGCAGCGCTGGGTGCGCTGCGCTGCACCAGCCACACGGGCCAGGCGTCGCACGAGTGCACGGAGGCCCTG		
	ORF Start: at 1		ORF Stop: end of

	sequence		
	SEQ ID NO: 388	622 aa	MW at 67376.2kD
NOV52b, CG51213-07 Protein Sequence	SINGAYWESI SHRPCPARGCTKTHREPGREHRALCSGTGSEPDIVWLPSRAGCPREEG RGASLSGHLGPQEVCSSELWCLSKSNRCITNSI PAAEGLTLCQTHITDKGWCYKRVCPVF GSRPEGVDGAWGPWTFWGDCSRCTCGGGVSSSSRHCDSPRPTIGGKYCLGERRRHRSN TDDCPPGSQDFREVQCSEFDSIPFRGKFYKWKTYRGGGVKACSLTCLAEGFNFYTERA AAVVDGTPCRPDTVDICVSGECKHVGCNDRVLGSDLREDKCRVCGDGSACETIEGVFS PASPGAGYEDVWVWIPKGSVHIFIQDLNLSLSHLALKGQESLLEGLPGTPQPHRLPL AGTTFQLRQGPDPQVSLEALGPINASLIVMLARTELPALRYRFNAPIARDSLPPYSW HYAPWTKCSAQCAGGSQVQAVECRNQLDSSAVAPHYCSAHSKLPKRQACNTEPCPD WVVGWNSLCSRSCDAGVRSRVSVCQRRVSAAEKALDDSAQPPRPVLEACHGPTCP PEWAALDWSECTPSCGPGLRHRVVLCKSADHRATLPPAHCSPAAKPPATMRCNLRRCP PARWVAGEWGECSAQCGVGQRQSVRCTSHTGQASHECTEAL		
	SEQ ID NO: 389	3199 bp	
NOV52c, CG51213-02 DNA Sequence	TAAAGGGTTT CAGCCTGGTGCCTGGTCCAGAGATAGTGGTGGTCATTGTTACCCCATAA TGGCATTGGTGCAAGTCCTTTCTTATCTATCTGTACGTGCCTCATAGCCATTATATA TAGGCAAGACAGGCATTAGGCTGCCATCTTGTAGATGAGTAAACTGAGGCCAGAGA GGGGAAATATATTGCAAGTTGGTAGCAGAATTGAGGTCTCTGCACAACTCAAATATGC CACAGTGCCTCTTGTGGAGAGGAGGACAAAAGCAGAGCTGAAATCATTTATCTTGAAG AGGTGTCAAGTGGGATTGCGACAGGACTGATGTGATATTTTAGATATGGCCAAGA GGACACAGTCTGAGTTTTAGCTGAGAAATGCTCTTATAAGGCAGAAAGCAGAGATT CTAGAGGACCTTTGAGGGAGAATGATTTGAGAACTCTTCCAGCTCTTACATAT GTACAGGTATCTCTCAGGGGTGACCTAGGAAGGGTCTCTTCTGTGGCCATTGATCG ATCCAGTCCCACATCTGGAAGCTTACAAGAATTGGGTTCAAAGCGGGGATTACACTT GATAATTACAGAAGGACCACTACTTCTTAGAGGAAAGACGCTGGGAGGTTGCTTAGG ATGTGGGCAAGAGGGTCAGAGAGGACCCTACTTTTATAGGAAAGACGCTGGGAG GTTGCTTAGGATGTGGGCCAAGAGGGTCAGAGATTTGCTTCACTGAACCTCACTGGG GCTTCTCCAGGGATATTAACCTGGACTTTAAGAGTCAGAGTGAGTCCCTGGGACTAGT TCAGCCCATCCAGGATTGAGACGGAAGAAGGTGGGGCTGATTTTCACTGGAGAAA GAGAGGCATGTCCACACAGACCTAACTCGGCATTGTCCCTCCCAACTCCCAACCC TCCACATAGCTTAAAGTGTGGGGGCTTCTCCAGTTAGATGGGGGAACAAAGAGAA CCAACAGCTGGAAAAAAGTAGAGATGAGGCCGTGGCCTAGTCATCATCAGGCCGAT TTCTCAGAACCACTTTCTCTCGGCTACTTTGCCCATCCCATAAAGAACCCCAA ATCCTTCTGTTCATTCTCAGCAGTTCACAGTTCTCTTCCAGAACTCAGAAGGCA CCAGGAAGTGAATTGCAAGTTCTGTAGAGCACAGACTCTGAATTAAGAGCTGGGTT AAATCCAGGCTATTCCCTTAGTAGCTGTGTGACCTTACCTGTCTGAAGCTTGGTTTT CTCCAGTAAGATGGGGTAGTACTGCCTAAAGAGGTATATGGCATGTATAAAGTGCTC CATAAATGAGCTTATTGGGAGAGTATAAGTCACAGGCCATGCCCGCAAGGGGATGC ACGAAGACCCACCGCAGCCAGGAAGGGAGCACCGGGCTCTCTGCTCTGGGACCGGCA GTGAGCCGGACATCTGGGTCTCCCAAGCCGGGCGGGCTGCCCGAGGAGGAAGGGAG GGGGGCGAGCCTGAGCGGGCACCTCGGCCCGCAGGAGGTCTGAGCGAGCTGTGGTGT CTGAGCAAGAGCAACCGGTGCATACCAACAGCATCCCGGCCCGGAGGGCACGCTGT GCCAGACGCACACCATCGACAAGGGGTGGTGCTACAAACGGGTCTGTGTCCCTTTGG GTCGCGCCAGAGGGTGTGGACGGAGCCTGGGGCCGTGGACTCCATGGGGCGACTGC AGCCGGACCTGTGGCGCGCGGTGCTCTTCTAGCCGTCACTGCGACAGCCCAAGGC CAACCATCGGGGGCAAGTACTGTCTGGGTGAGAGAAGCGGCACCGCTCTCGCAACAC GGATGACTGTCCCTTGGCTCCAGGACTTCAGAGAAGTGCAAGTGTCTGAATTTGAC AGCATCCCTTTCCGTGGGAAATTCTACAAGTGGAAAACGTACCGGGAGGGGGCGTGA AGGCCTGCTCGCTCACGTGCCTAGCGGAAGGCTTCAACTTCTACACGGAGAGGGCGGC AGCCGTGGTGGACGGGACACCTGCGGTCCAGACACGGTGGACATTTGCGTCAGTGGC GAATGCAAGCACGTGGGCTGCGACCGAGTCTGGGCTCCGACTGCGGGAGGACAAGT GCCGAGTGTGTGGCGGTGACGGCAGTGCCTGCGAGACCATCGAGGGCGTCTTACGCC AGCCTCACCTGGGGCCGGGTACGAGGATGTCTGTGGATTCCCAAGGCTCCGTCCAC ATCTTCATCCAGGATCTGAACCTCTCTCTCAGTCACTTGGCCCTGAAGGGAGACCAGG AGTCCCTGCTGCTGGAGGGGTGCCCGGGACCCCCAGCCCCACCGTCTGCTCTAGC TGGGACCACTTTCAACTGCGACAGGGGCCAGACCAGGTCCAGAGCCTCGAAGCCCTG GGACCGATTAAATGATCTCTCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG TCCGCTACCGCTTCAATGCCCCATCGCCCGTGAATCGCTGCTGCTGCTGCTGCTGCTG CTATGCGCCCTGGACCAAGTGCTCGGCCAGTGTGAGGCGGTAGCCAGGTGACGGCG		

	GTGGAGTGCCGCAACCAGCTGGACAGCTCCGCGGTGCGCCCCCACTACTGCAGTGCCC ACAGCAAGCTGCCAAAAGGCAGCGCGCCTGCAACACGGAGCCTTGCCCTCCAGACTG GGTTGTAGGGAACCTGGTCGCTCTGCAGCCGAGCTGCGATGCAGGCGTGCGCAGCCGC TCGGTCGTGTGCCAGCGCCGCGTCTCTGCCGCGGAGGAGAAGGCGCTGGACGACAGCG CATGCCCGCAGCCGCGCCCACTGTACTGGAGGCCCTGCCACGGCCCCACTTGCCCTCC GGAGTGGGCGCCCTCGACTGGTCTGAGTGCACCCAGCTGCGGGCCGGGCTCCGC CACCGCGTGGTCTCTTTGCAAGAGCGCAGACCACCGCGCCACGCTGCCCCGGCGCACT GCTCACCCGCGCCAAGCCACCGGCCACCATGCGCTGCAACTTGCGCCGCTGCCCCC GGCCCGCTGGGTGGCTGGCGAGTGGGGTGAAGTGTCTGACAGTGCGGCGCTCGGGCAG CGGCAGCGCTCGGTGCGCTGCACCAGCCACACGGGCCAGGCGTGCACAGAGTGCACGG AGGCCCTGC		
	ORF Start: at 1297		ORF Stop: at 3199
	SEQ ID NO: 390	634 aa	MW at 68853.1kD
NOV52c, CG51213-02 Protein Sequence	YCLKRYMACIKCSINGAYWESISHRPCFARGCTKTHREPGREHRALCSGTGSEPDIVV LPSRAGCPREEBGRGASLSGHLGPQEVCSLWCLSKSNRCITNSIPAAEGTLCQHTIID KGWCYKRVCPVFGSRPEGVDGAWGPWTPWGDSCRTCGGVSSSRHCDSPRPTIGGKY CLGERRRHRSCTDDCPPGSQDFREVQCSEFDSIPFRGKFKWKTYRGGGVKACSLTC LAEGFNFYTERAAAVVDGTPCRPDTVDICVSGECKHVGCIDRVLGSDLREDKRCVCGGD GSACETIEGVFSPASPGAGYEDVWVWIPKGSVHIFIQDLNLSLSHLALKGDQESLLEGL LPGTPQPHRLPLAGTTTFLRQGPQVQSLEALGPINASLIVMVLARTELPALRYRFNA PIARDSLPYSPWHYAPWTKCSAQCGGSQVQAVECRNQLDSSAVAPHYCSAHSKLPKR QRACNTEPCPPDWVGNWSLCSRSCDAGVRSRSVVCQRRVSAAEEKALDDSAACQPRP PVLEACHGPTCPPEWAALDWSECTPSCGPGLRHRVVLCKSADHRATLPPAHCSPAAPK PATMRCNLRRCPPARWVAGEWGECSAQCGVGQRQSVRCTSHGTQASHECTEAL		
	SEQ ID NO: 391	3700 bp	
NOV52d, CG51213-03 DNA Sequence	CTGACATTCCACCCTTGACACCCCCAACATCCTAACTTAGCTGGTAACTGCAGCACC CTCTAAGGAATTCTTAAAGAATTCTGAAGCTACTCCTCAACATCTGCTGTGACCCAGG TATCCTAACAATGATCATGGTGTCTGACATTTACTGAGCTCTCACTATGGGCTAAGCA TGTGCTGTGTGTCAACATCTAACTCCTGACAATCCTGTAGCCCCACGTTACAGAG GAAGGGACTGAGCCATAGCATAGGGAGGATGACTTGTCCAAGGCCACAGTTTGAGACC ATGACAGAGCTGGGATTTAAATCCAGGTCTCTCATGACTCTCTAAATTTTACAAAGGG GCAGGGGAGGGGAGGAGCTGTCAAAATATCAAGCTTGGGCTGGCACTGGCTATATGTT GAATTGAGCCTTCCTTTTAGTTTTGAAGGAACATCTTTCAGGCCATCTTGGCAAAGG GGGATTTATTTACTAAATGTGAAGTGGTTAATATATGTAAAGGGTTCAGCCTGGTGCC TGGTCCAGAGATAGTGGTGGTCATTGTTACCCCATATAGGCATTGGTGCAAGTCCTTT CTTATCTATCCTGTCACTGCCCTCATAGCCATTATATAGGCAAGACAGGCATTAGGC TGCCCATCTTGTAGATGAGTAACTGAGGCCAGAGAGGGGAAATATATTGCAAGTTG GTAGCAGAATTGAGGTCTCTGCACAACCTCAAATATGCCACAGTGCCTCCTTGTGGAGA GGAGGACAAAAGCAGAGCTGAAATCATTATCTTGAAGAGGTGTGAGAAGTGGGATTGC GACAGGACTGATGTGATATTTTTAGATATGGCCAAGAGGACACAGTCTGAGTTTTTAG CTGAGAAATGTCTCTATAAGGCAGAAGGCAGAGATTTAGAGGACCTTTGAGGGAGA ATGTATTTGAGAACTACTTTCAGCTTCTTACATATGTACAGGTATCTCTCAGGGGC TGACCTAGGAAGGGTCCTTTCTGTGGCCATTGATCGATCCAGTCCCACATCTGGA GCTTACAAGAATTGGGTTCAAAGCGGGGATTACACTTGATAATTACAGAAGGACCACC TACTTCTTAGAGGAAAGACGCTGGGAGGTTGCTTAGGATGTGGGCCAAGAGGGTCAGA GAGGACCACCTACTTTTTAGAGGAAAGACGCTGGGAGGTTGCTTAGGATGTGGGCCAA GAGGGTCAGAGATTTTGCTTCACTGAACTCACTGGGGCTTCTCCAGGGATATTAACC TGGACTTTAAGAGTCAGAGTGAGTCCCTGGGACTAGTTCAGCCCATCCAGGATTGAGA CGGGAAGAAGGTGGGGCTGATTTTTCACCTGGAGAAAGAGAGGCATGTCCACACAGA CCTAACTCGGCATTGTCCCTCCCAAACCTCCACCCCTCCACATAGCTTAAAGTGTT GGGGGCTTCTCCAGTTTAGATGGGGGAACAAAGAGAACCAACAGCTGGA AGATAGGCGCTTGGCCTAGTCAATCCAGGCCGATTCTCAGAACCCACTTTCT CTTCGGCTACTTTGCCCATCCATAAAGAACCCAAATCCTTCTGTTCATTCTCTCA GCAGTTCACACGTTTCTTCCAGAACTCAGAAGGCACAGGAAGTGAATTGCAAGT TCGTTAGAGCACAGACTCTGAATTAAAGAGCTGGGTTAACTCCAGGCTATTCCCTTA GTAGCTGTGTGACCTTACCTGTCTGAAGCTTGGTTTTCTCCAGTAAGATGGGGTAGT ACTGCCTAAAGAGGTATATGGCATGTATAAGTGCTCCATAAATGAGCTTATTGGGA GAGTATAAGTTCACAGGCCATGCCCGCAAGGGGATGCACGAAGACCCACCGCAGGCC		

	GGAAGGGAGCACGGGGCTCTCTGCTCTGGGACCGGCAGTGAGCCGGACATCTGGGTCC TCCAAGCCGGGCGGGCTGCCCCAGGGAGGAAGGGAGGGGGCGAGCCTGAGCGGGCA CCTCGGCCCGCAGGAGGTCTGCAGCGAGCTGTGGTGTCTGAGCAAGAGCAACCGGTGC ATCACCAACAGCATCCCGGCCGCCGAGGGCACGCTGTGCCAGACGCACACCATCGACA AGGGGTGGTGCTACAAACGGGTCTGTGTCCCTTTGGGTGCGCCCCAGAGGGTGTGGA CGGAGCCTGGGGGGCGTGGACTCCATGGGGCGACTGCAGCCGACCTGTGGCGGGC GTGTCTCTTCTAGCCGTCACTGCGACAGCCCCAGGCCAACCATCGGGGGCAAGTACT GTCTGGGTGAGAGAAGGCGGCACCGCTCCTGCAACACGGATGACTGTCCCCCTGGCTC CCAGGACTTCAGAGAAGTGAGTGTCTGAATTTGACAGCATCCCTTTCCGTGGGAAA TTCTACAAGTGAAAACGTACCGGGGAGGGGGCGTGAAGGCCTGCTCGCTCACGTGCC TAGCGGAAGGCTTCAACTTCTACACGGAGAGGGCGGCAGCCGTGGTGGACGGGACACC CTGCCGTCCAGACACGGTGGACATTTGCGTCAGTGGCGAATGCAAGCACGTGGGCTGC GACCGAGTCTTGGGCTCCGACCTGCGGGAGGACAAGTGCCGAGTGTGTGGCGGTGACG GCAGTGCTGCGAGACCATCGAGGGCGTCTTACGCCCAGCCCTACCTGGGGCCGGGTA CGAGGATGTCGTCTGGATTCCCAAAGGCTCCGTCCACATCTTTCATCCAGGATCTGAAC CTCTCTCTCAGTCACTTGGCCCTGAAGGGAGACCAGGAGTCCCTGCTGCTGGAGGGC TGCCCGGAGCCCCCAGCCCCACCGTCTGCCCTTAGCTGGGACCACTTTCAACTGCG ACAGGGGCCAGACCAGGTCCAGAGCCTCGAAGCCCTGGGACCGATTAAATGCATCTCTC ATCGTCATGGTGCTGGCCCGGACCGAGCTGCCCTGCCCTCCGTACCGCTTCAATGCC CCATCGCCCGTGACTCGCTGCCCCCTACTCCTGGCACTATGCGCCCTGGACCAAGTG CTCGGCCCAGTGTGCAGGCGGTAGCCAGGTGCAGGCGGTGGAGTGCCGCAACACAGTG GACAGCTCCGCGGTGCCCCCCTACTGCACTGCCACAGCAAGCTGCCCAAAAGGC AGCGCGCTGCAACACGGAGCCTTGCCCTCCAGACTGGGTGTAGGGAAGTGGTGCCT CTGCAGCCGAGCTGCGATGCAGGCGTGCAGCCGCTCGGTGCTGTGCCAGCGCCGC GTCTCTGCCGCGGAGGAGAAGGCGCTGGACGACAGCGCATGCCCGCAGCCCGCCCCAC CTGTACTGGAGGCTGCCACGGCCCCACTTGCCCTCCGGAGTGGGGCGGCCCTCGACTG GTCTGAGTGACCCCCAGCTGCGGGCCGGGCCCTCCGCCACCGCTGGTCTTTGCAAG AGCGCAGACCACCGCGCCACGCTGCCCCCGGCGCACTGCTCACCCGCGGCCAAGCCAC CGGCCACCATGCGCTGCAACTTGCGCCGCTGCCCCCGGCCCGCTGGGTGGCTGGCGA GTGGGGTGAGTGCTTGCACAGTGGCGCTGCGGCAGCGGCAGCGCTCGGTGCGCTGC ACCAGCCACACGGGCCAGGCGTCGCACGAGTGACGGAGGCCCTGC		
	ORF Start: at 1798		ORF Stop: at 3700
	SEQ ID NO: 392	634 aa	MW at 68754.0kD
NOV52d, CG51213-03 Protein Sequence	YCLKRYMACIKCSINGAYWESISHRPPARGCTKTHREPGRHGHALCSGTGSEPDIVV LPSRAGCPREEGRGASLSGHLGPQEVCSSELWCLSKSNRCITNSIPAAEGTLCTHTID KGWCYKRVCPVFGSRPEGVDGAWGPWTPWGDCSRTCGGGVSSSRHCHDSPRPTIGGKY CLGERRRRHSCNTDDCPPGSQDFREVQCSEFDSIPFRGKFYKWKTYRGGGVKACSLTC LAEGFNFYTERAAAVVDGTPCRPDTVDICVSGECKHVGCNDRVLGSDLRECKRCVCGD GSACETIEGVFSPASPGAGYEDVWVWIPKGSVHIFIQDLNLSLSHLALKGQDESLLLEG LPGTTPQPHRLPLAGTTFQLRQPDQVQSLEALGPINASLIMVLARTELPALRYRFNA PIARDSLPPYSWHYAPWTKCSAQAGGSQVQAVECRNQLDSSAVAPHYCSAHSKLPKR QRACNTEPCPPDWVGNWSLCSRSCDAGVRSRVSVCQRRVSAAEEKALDDSAACPQPRP PVLEACHGPTCPPEWAALDWSECTPSCGGLRHRVVLCKSADHRATLPPAHCSPAAKP PATMRCNLRRCPPARWVAGEWGECSAQCGVGQRQSVRCTSHTGQASHECTEAL		
	SEQ ID NO: 393	2804 bp	
NOV52e, CG51213-04 DNA Sequence	TGGCCAGCCAGGCTGAAGCGATCGGTGAGCCGAGAGCGCTACGTGGAGACCCTGGTG GTGGCTGACAAGATGATGGTGGCTATCACGGGCGCCGGGATGTGGAGCAGTATGTCC TGGCCATCATGAACATTCAGGTTGCCAACTTTTCCAGGACTCGAGTCTGGGAAGCAC CGTTAACATCCTCGTAACCTCGCTCATCCTGCTCACGGAGGACCAGCCCACTCTGGAG ATCACCCACCATGCCGGAAGTCCCTGGACAGCTTCTGTAAGTGGCAGAAATCCATCG TGAACCACAGCGGCATGGCAATGCCATTCCAGAGAACGGTGTGGCTAACCATGACAC AGCAGTGCTCATCACAGCTATGACATCTGCATCTACAAGAACAACCTGCGGCACA CTAGGCTTGGCCCGGTGGGCGGAATGTGTGAGCGCGAGAGAAGCTGCAGCGCTCAATG AGGACATTGGCCTGGCCACAGCCTTACCATTGCCACAGAGATCGGGCACACATTGGG CATGAACCATGACGGCGTGGGAAACAGCTGTGGGGCCCGTGGTCAAGACCCAGCCAAG CTCATGGCTGCCACATTACCATGAAGACCAACCCATTCTGTGGTTCATCTGCAGCC GTGACTACATCACCAGCTTTCTAGACTCGGGCTGGGGCTCTGCCTGAACAACCGGCC CCCCAGACAGGACTTTGTGTACCCGACAGTGGCACCGGGCCAAGCCTACGATGCAGAT		

	GAGCAATGCCGCTTTCAGCATGGAGTCAAATCGCGTCAGTGTAATACGGGGAGGTCT GCAGCGAGCTGTGGTGTCTGAGCAAGAGCAACCGGTGCATCACCACAGCATCCCGGC CGCCGAGGGCACGCTGTGCCAGACGCACACCATCGACAAGGGGTGGTCTACAAACGG GTCTGTGTCCCTTTGGGTGCGCCCCAGAGGGTGTGGACGGAGCCTGGGGGCCGTGGA CTCCATGGGGCGACTGCAGCCGACCTGTGGCGGGCGGTGTCTCTTCTAGCCGTCA CTGCGACAGCCCCAGGCCAACCATCGGGGGCAAGTACTGTCTGAGTGAGAGAAGGCGG CACCCTCTCTGCAACACGGATGACTGTCCCTTGGCTCCAGGACTTCAGAGAAGTGC AGTGTCTGAATTGACAGCATCCCTTTCCGTGGGAAATTCTACAAGTGAAAACGTA CCGGGAGGGGGCGTGAAGGCTGCTCGCTCAGAGCCTAGCGGAAGGCTTCAACTTC TACACGAGAGGGCGGCAGCCGTGGTGGACGGGACACCTGCGCTCCAGACACGGTGG ACATTTGCGTCAGTGGCGAATGCAAGCACGTGGGCTGCGACCGAGTCTGGGCTCCGA CCTGCGGGAGGACAAGTGCCGAGTGTGTGGCGGTGACGGCAGTGCCCTGCGAGACCATC GAGGGCGTCTTCAGCCAGCCTCACCTGGGGCCGGGTACGAGGATGTCGTCTGGATTCT CCAAAGGCTCCGTCCACATCTTCATCCAGGATCTGAACCTCTCTCTCAGTCACTTGGC CCTGAAGGGAGACCAGGAGTCCCTGCTGCTGGAGGGGTGCTGGGACCCCCAGCCC CACCCTCTGCTCTAGCTGGGACCACTTTCAACTGCGACAGGGGCCAGACCAGGTCC AGAGCCTCGAAGCCTGGGACCGATTAAATGCATCTCTCATCGTCATGGTGTGGCCCG GACCGAGCTGCTGCTCCCTCGCTACCGCTTCAATGCCCCCATCGCCGCTGACTCGCTG CCCCCTACTCTGGCACTATGCGCCCTGGACCAAGTGTCTGGCCCACTGTGCAGGCG GTAGCCAGGTGCAGGCGGTGGAGTGC CGCAACCAGCTGGACAGCTCCGCGGTGCGCCC CCACTACTGCACTGCCACAGCAAGCTGCCCAAAGGCAGCGCGCTGCAACACGGAG CCTTGCCCTCCAGACTGGGTGTAGGGAACTGGTCTGCTGTCAGCCGAGCTGCGATG CAGGCGTGCAGTGTGCTCGTGTGTCAGCGCGCGCTCTGCTGCCGCGAGGAGAA GGCGCTGGACGACAGCGCATGCCCGCAGCGCGCCACCTGTACTGGAGGCTGCCAC GGCCCCACTTGCCCTCCGAGTGGGCGGCCCTCGACTGGTCTGAGTGACCCCCAGCT GCGGGCCGGGCTCCGCCACCGCTGGTCTTTGCAAGAGCGCAGACCACCGCGCCAC GCTGCCCCCGGCGCACTGCTCACCCGCGCCAAGCCACCGGCCACCATGCGCTGCAAC TTGCGCGCTGCCCCCGGCCGCTGGGTGGCTGGCGAGTGGGGTGAGTGCTCTGCAAC AGTGGCGGTGCGGCGAGCGGCGAGCGCTCGGTGCGCTGCACAGCCACAGGGCCAGGC GTCGACGAGTGACGAGGCGCTGCGGCCGCCACCACGCGAGTGTGAGGCCAAG TGCGACAGCCCAACCCCCGGGACGGCCCTGAAGAGTGCAAGGATGTGAACAAGGTG CCTACTGCCCCCTGGTGTCAAATTTAGTTCTGACGCGAGCCTACTTCCGCCAGAT GTGCTGCAAAACCTGCCAGGGCCACTAGGGGGCGCGCGGCCACCCGAGCCACAGTGG CGGGGTCTCCGCCGCCAGCCCTGCAGCGGGCGGCCAAAGGGGGCCCCGGGGGGCGG GAACTGGGAGGGAAGGGTGAGACGGAGCCGGAAGTTATTTATTGGGAACCCCTGCAGG GCCCTGGCTGGGGGGATGGA		
	ORF Start: ATG at 71	ORF Stop: TAG at 2636	
	SEQ ID NO: 394	855 aa	MW at 93285.7kD
NOV52e, CG51213-04 Protein Sequence	MMVAYHGRRDVEQVLAIMNIQVAKLFQDSSLGSTVNI LVTRLILLTEDQPTLEITHH AGKSLDSFCKWQKSI VNHSGHNAI PENG VANHDTAVLITRYDICIYKNKPCGTLGLA PVGGM CERERSCSVNEDIGLATAFTIAHEIGHTFGMNH DGVGNSCGARGQDPAKLMAA HITMKTNPFWSSCSRDIYITSLD SGLGLCLNNRPPRQDFVYPTVAPQAYDADEQCR FQHG VKS RQCKYGEVCS ELWCLSKSNRCITNSIPAAEGTLCQHTIDKGWCYKRVCPV FGSRPEGVDGAWGPWTPWGD CSR TC GGVSSSRHCDSPRPTIGGKYCLSERRRHRSC NTDDCPGSGQDFREVQCSEFDSI PFRGKFYKWKTYRGGGVKACSLTSLAEGFNFYTER AAAVVDGTPCRPDTVDICVSGECKHVGCDRVLGSDLREDKCRVCGGDGSACETIEGVF SPASPGAGYEDVWVWIPKGSVHIFIQDLNLSLSHLALKGDQESLLEGLPGTPQPHRLP LAGTTFQLRQGPDPVQSLEALGPINASLIVMLARTELPALRYRFNAPIARDSLPPYS WHYAPWTKCSAQ CAGGSQVQAVECRNQLDSSAVAPHYCSAHSKLPKRQRANTPECP DWVGNWLSLCSRCDAGVRSRVVCQRRVSAAEKALDDSA CPQPRPVLEACHGPTC PPEWALDWSECTPSCGPGLRHRVVLCKSADHRATLPPAHCS PAAKPPATMRCNLRR PPARWVAGEWGECSAQCGVGQRQSVRCTSH TGQASHECTEALRPPTTQQCEAKCDSP TPGDGPEECKDVNKVAYCPLVLKFQFCRAYFRQMCCKTCQGH		
	SEQ ID NO: 395	3400 bp	
NOV52f, CG51213-05 DNA Sequence	CGGTCTCAAGATGAGTTCTGTCCAGTCTGGAGAGCTATGAGATCGCCTTCCCCACCC GCGTGGACCACAACGGGGCACTGCTGGCCTTCTCGCCACCTCCTCCCCGGAGGCAGCG CCGCGGCACGGGGCCACAGCCGAGTCCCGCTCTTCTACAAAGTAGCTCGCCACG ACCCACTTCTGTGAACCTGACCCGAGCTCCCGTCTACTGGCAGGGCACGTCTCCG		

	<p>TGGAGTACTGGACACGGGAGGGCCTGGCCTGGCAGAGGGCGGCCCGGCCCACTGCCT  CTACGCTGGTCACTGCAGGGCCAGGCCAGCAGCTCCCATGTGGCCATCAGCACCTGT  GGAGGCCCTGCACGGCCTGATCGTGGCAGACGAGGAAGAGTACCTGATTGAGCCCTGTC  ACGGTGGGCCCAAGGGTTCTCGGAGCCCGGAGGAAAGTGGACCACATGTGGTGTACAA  GCGTTCCCTCTCTGCGTCACCCCCACCTGGACACAGCCTGTGGAGTGAGAGATGAGAAA  CCGTGGAAAGGGCGGCCATGGTGGCTGCGGACCTTGAAGCCACCGCCTGCCAGACCCC  TGGGGAATGAAACAGAGCGTGGCCAGCCAGGCCTGAAGCGATCGGTGAGCCGAGAGCG  CTACGTGGAGACCTTGGTGGTGGCTGACAAGATGATGGTGGCTATCACGGGCGCCGG  GATGTGGAGCAGTATGTCTGGCCATCATGAACATTGTTGCCAAACTTTTCCAGGACT  CGAGTCTGGGAAGCACCGTTAACATCCTCGTAACTCGCCTCATCCTGCTCAGGAGGA  CCAGCCCACTCTGGAGATCACCCACCATGCCGGGAAGTCCCTAGACAGCTTCTGTAAG  TGGCAGAAATCCATCGTGAACCACAGCGGCCATGGCAATGCCATTCCAGAGAACGGTG  TGGCTAACCATGACACAGCAGTGTCTCATCACACGCTATGACATCTGCATCTACAAGAA  CAAACCTTGCGGCACACTAGGCCCTGGCCCCGGTGGGCGGAATGTGTGAGCGCGAGAGA  AGCTGCAGCGTCAATGAGGACATTGGCCTGCCACAAGCGTTCACCATTGCCACGAGA  TCGGGCACACATTTCGGCATGAACCATGACGGCGTGGGAAACAGCTGTGGGGCCCGTGG  TCAGGACCCAGCCAAGTCTATGGCTGCCACATTACCATGAAGACCAACCCATTCTGT  TGGTCATCCTGCAACCGTGACTACATCACAGCTTCTAGACTCGGGCCTGGGGCTCT  GCCTGAACAACCGGCCCCCCAGACAGGACTTTGTGTACCCGACAGTGGCACCGGGCCA  AGCCTACGATGCAGATGAGCAATGCCGCTTTCAGCATGGAGTCAAATCGCGTCAGTGT  AAATACGGGGAGGTCTGCAGCGAGCTGTGGTGTCTGAGCAAGAGCAACCGTGCATCA  CCAACAGCATCCCGCCCGGAGGGCAGCTGTGCCAGACGCACACCATCGACAAGGG  GTGGTGCTACAAACGGGTCTGTGTCCCTTTGGGTGCGGCCAGAGGGTGTGGACGGA  GCCTGGGGCCGTGGACTCCATGGGGCGACTGCAGCCGACCTGTGGCGCGCGCTGT  CCTCTTCTAGTCTGCTACTGCGACAGCCCCAGGCCAACCATCGGGGGCAAGTACTGTCT  GGGTGAGAGAAGGCGGCACCGCTCTGCAACACGGATGACTGTCCCCCTGGCTCCAG  GACTTCAGAGAAGTGCAGTGTCTGAATTTGACAGCATCCCTTTCGGTGGGAAATTCT  ACAAGTGGAAAACGTACCGGGGAGGGGGCGTGAAGGCCTGCTCGCTCAGAGCCTAGC  GGAAGGCTTCAACTTCTACACGGAGAGGGCGGCAGCCGTGGTGGACGGGACACCCCTGC  CGTCCAGACACGGTGGACATTTGCGTCAAGTGGCGAATGCAAGCAGTGGGTGCGACC  GAGTCTGGGCTCCGACCTGCGGGAGGACAAGTGCCGAGTGTGTGGCGGTGACGGCAG  TGCCTGCGAGACCATCGAGGGCGTCTTCAGCCCAGCCTCACCTGGGGCCCGGTACGAG  GATGTCTGTCTGGATTCCCAAAGGCTCCGTCCACATCTTATCCAGGATCTGAACCTCT  CTCTCAGTCACTTGGCCCTGAAGGGAGACCAGGAGTCCCTGTCTGGAGGGGCTGCC  TGGGACCCCCAGCCCCACCGTCTGCCTCTAGCTGGGACCACCTTTCAACTGCGACAG  GGGCCAGACCAGGTCCAGAGCCTCGAAGCCCTGGGACCGATTAATGCATCTCTCATCG  TCATGGTGTGGCCCCGACCGAGCTGCCTGCCCTCCGCTACCGCTTCAATGCCCCCAT  CGCCCGTGACTCGCTGCCCCCTACTCCTGGCACTATGCGCCCTGGACCAAGTGCTCG  GCCAGTGTGCAAGCGGTAGCCAGGTGCAAGCGGTGGAGTGCAGCAACCAAGCTGTGACA  GCTCCGCGGTGCGCCCCCACTACTGCAGTGCCACAGCAAGCTGCCCAAAGGCAGCG  CGCCTGCAACACGGAGCCTTGCCCTCCAGACTGGGTTGTAGGGAAGTGGTCTCTGTC  AGCCGAGCTGCGATGCAGGCGTGCAGTGTGCTCGGTCTGTGCCAGCGCCGCGTCT  CTGCCGCGGAGGAGAAGGCGCTGGACGACAGCGCATGCCCGCAGCCGCGCCACCTGT  ACTGGAGGCCTGCCACGGCCCCACTTGCCCTCCGAGTGGGCGGCCCTCGACTGGTCT  GAGTGCACCCCCAGCTGCGGGCCGGGCTCCGCCACCGCGTGGTCTTTGCAAGAGCG  CAGACCACCGCGCCACGCTGCCCCCGGCGCACTGCTCACCCGCGCCCAAGCCACCGGC  CACCATGCGCTGCAACTTGCGCCGCTGCCCCCGGCCCGCTGGGTGGCTGGCGAGTGG  GGTGAAGTCTCTGCACAGTGCAGCGTGGGCGAGCGGACGCTCGGTGCGCTGCACCA  GCCACACGGGCCAGGCGTGCACGAGTGCACGAGGCCCTGCGGCCGCCACACGCA  GCAGTGTGAGGCCAAGTGCAGACAGCCCCAACCCCGGGGACGCCCTGAAGAGTGAAG  GATGTGAACAAGGTGCGCTACTGCCCCCTGGTGTCAAATTTCAAGTTCTGCAGCCGAG  CCTACTTCCGCCAGATGTGCTGCAAAACCTGCCAGGGCCACTAGGGGGCGCGGGCAC  CCGGAGCCACAGCTGGCGGGGTCTCCGCGCCAGCCCTGCAGCGGGCCGCCAAAGGG  GGCCCCGGGGGGGGGGAAGTGGGAGGGAAGGTTGAGACGGAGCCGGAAGTTATTAT  TGGGAACCCCTGCAGGGCCCTGGCTGGGGGATGGA</p>
	<p>ORF Start: at 1</p>
	<p>ORF Stop: TAG at 3232</p>
NOV52f, CG51213-05.	<p>SEQ ID NO: 396   1077 aa   MW at 118071.4kD</p> <p>RSQDEFLSSLESYETAFPTVRVDHNGALLAFSPPPRRRQRRGTGATAESRLFYKVASPS  THFLNLTRSSRLLAGHVSVEYWTREGLAQRAARPHCLYAGHLQGGQASSSHVAISTC</p>

Protein Sequence	GGLHGLIVADEEYLI EPLHGGPKGSRSP EESGPHVVYKRSSLRHPHLDTACGVRDEK PWKGRPWLLRTLKPPPARPLGNETERGQPLKRSVSRERYVETLVVADKMMVAYHGRR DVEQYVLAIMNIVAKLFQDSSLGSTVNILVTRLILLTEDQPTLEITHHAGKSLDSFCK WQKSI VNHSGHGNAI PENGVANHDTAVLITRYDICIYKNKPCGTGLGLAPVGGMCERER SCSVNEDIGLPQAFITIAHEIGHTFGMNH DGVGNSCGARGQDPAKLMAAHITMKTNP FV WSSCNRDYITSFLDSGLGLCLNNRPPRQDFVYPTVAPGQAYDADEQCRFQHGVKSRQC KYGEVCSSELWCLSKSNRCITNSI PAAEGTLCQHTTIDKGWCYKRVCPVFGSRPEGVDG AWGPWTPWGDCSRTC GGGVSSSRHCDSPRPTIGGKYCLGERRRHRSCNTDDCPPGSQ DFREVQCSEFDSI PFRGKFYKWKTYRGGGVKACSLTSLAEGFNFYTERAAAVVDGTPC RPDITVDICVSGECKHVGCDRVLGSDLREDKCRVCGGDGSACETIEGVFSPASPGAGYE DVVWI PKGSVHIFIQDLNLSLSHLALKG DQESLLLEGLPGTPQPHRLPLAGTTFQLRQ GPDQVQSLEALGFINASLIWMVLARTELPALRYRFNAPIADNSLPYPYSHWYAPWTKCS AQCAGGSQVQAVECRNQLDSSAVAPHYCSAHSKLPKRQRCANTBPCPPDWVGNWNSLC SRSCDAGVRSRVSVCQRRVSAAEEKALDDSACQP RPVLEACHGPTCPPEWAALDWS ECTPSCGPGLRHRVVLCKSADHRATLPPAHCSPAAKPATMRCNLRRCP PARVWAGEW GECSAQCGVGQRQRSVRCTSH TGGQASHECTEALRPPTTQQCEAKCDSPTPGDGPEECK DVNKVAYCPLVLKFQFCRAYFRQMCCKTCQGH		
	SEQ ID NO: 397		978 bp
NOV52g, CG51213-06 DNA Sequence	TCCATAAATGGAGCTTATTGGGAGAGTATAAGTCACAGGCCATGCCCGCAAGGGGAT GCACGAAGACCCACCGCAGCCAGGAAGGGAGCACC GGCTCTCTGCTCTGGGACCGG CAGTGAGCCGGACATCTGGTCTCTCC AAGCCGGCGGGCTGCCCCAGGGAGGAAGGG AGGGGGGCGAGCCTGAGCGGGCACCTCGGCCCGCAGGAGGTCTGCAGCGAGCTGTGGT GTCTGAGCAAGAGCAACCGGTGCATCACCAACAGCATCCCGGCCGCCGAGGGCACGCT GTGCCAGACGCACACCATCGACAAGGGGTGGTGCTACAAACGGGTCTGTGTCCCTTT GGGTCCGCGCCAGAGGGTGTGGACGGAGCCTGGGGGCCGTGGACTCCATGGGGCGACT GCAGCCGGACCTGTGGCGGGCGGCTGTCTCTTCTAGCCGTCACTGCGACAGCCCCAG GCCAACCATCGGGGGCAAGTACTGTCTGGGTGAGAGAAGGCGGCACCGCTCCTGCAAC ACGGATGACTGTCCCCCTGGCTCCAGGACTTCAGAGAAGTGCAGTGTCTGAATTG ACAGCATCCCTTTCCGTGGGAAATTCTACAAGTGGAAAACGTACCGGGGAGGGGGCGT GAAGGCCTGCTCGCTACGTGCCTAGCGGAAGGCTTCAACTTCTACACGGAGAGGGCG GCAGCCGTGGTGGACGGGACACCCTGCCGTCCAGACACGGTGGACATTTGCGTCAGTG GCCAATGCAAGCACGTGGGCTGCGACCGAGTCTTGGGCTCCGACCTGCGGGAGGACAA GTGCCGAGTGTGTGGCGGTGACGGCAGTGCTGCGAGACCATCGAGGGCGTCTTCAGC CCAGCCTCACCTGGGGCCGGGTACGAGGATGTCTGTGGATTCCCAAAGGCTCCGTCC ACATCTTCATCCAGGATCTGAACCTCTCTCTCAGTCACTTGGCCCTGAAG		
	ORF Start: at 1		ORF Stop: end of sequence
	SEQ ID NO: 398	326 aa	MW at 35330.2kD
NOV52g, CG51213-06 Protein Sequence	SINGAYWESI SHRPCARGCTKTHREPGREHRLCSGTGSEPDIVWLP SRAGCPREEG RGASLSGHLGPQEVCSSELWCLSKSNRCITNSI PAAEGTLCQHTTIDKGWCYKRVCPF GSRPEGVDGAWGPWTPWGDCSRTC GGGVSSSRHCDSPRPTIGGKYCLGERRRHRSCN TDDCPPGSQDFREVQCSEFDSI PFRGKFYKWKTYRGGGVKACSLTCLAEGFNFYTERA AAVVDGTPCRPDTVDICVSGECKHVGCDRVLGSDLREDKCRVCGGDGSACETIEGVFS PASPGAGYEDVVWI PKGSVHIFIQDLNLSLSHLALK		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 52B.

Table 52B. Comparison of NOV52a against NOV52b through NOV52g.		
Protein Sequence	NOV52a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV52b	54..465 211..622	412/412 (100%) 412/412 (100%)

NOV52c	54..465 223..634	412/412 (100%) 412/412 (100%)
NOV52d	54..465 223..634	412/412 (100%) 412/412 (100%)
NOV52e	54..523 386..855	469/470 (99%) 469/470 (99%)
NOV52f	54..523 608..1077	469/470 (99%) 469/470 (99%)
NOV52g	54..169 211..326	116/116 (100%) 116/116 (100%)

Further analysis of the NOV52a protein yielded the following properties shown in Table 52C.

Table 52C. Protein Sequence Properties NOV52a	
PSort analysis:	0.6400 probability located in plasma membrane; 0.5231 probability located in outside; 0.1900 probability located in lysosome (lumen); 0.1000 probability located in endoplasmic reticulum (membrane)
SignalP analysis:	Cleavage site between residues 37 and 38

5. A search of the NOV52a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 52D.

Table 52D. Geneseq Results for NOV52a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV52a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAU01292	Human Thrombospondin repeat domain protein 2, TSR2 - Homo sapiens, 523 aa. [WO200123561-A2, 05-APR-2001]	1..523 1..523	523/523 (100%) 523/523 (100%)	0.0
AAU97888	Human aggrecanase protein #2 - Homo sapiens, 1104 aa. [WO200234895-A2, 02-MAY-2002]	54..523 634..1103	470/470 (100%) 470/470 (100%)	0.0
AAU72890	Human metalloprotease partial protein sequence #2 - Homo sapiens, 1103 aa. [WO200183782-A2, 08-NOV-2001]	54..523 634..1103	470/470 (100%) 470/470 (100%)	0.0

AAB74945	Human ADAM type metal protease MDTs2 protein SEQ ID NO:10 - Homo sapiens, 1103 aa. [JP2001008687-A, 16-JAN-2001]	54..523 634..1103	470/470 (100%) 470/470 (100%)	0.0
AAB72300	Human ADAMTS-10 alternative amino acid sequence - Homo sapiens, 1072 aa. [WO200111074-A2, 15-FEB-2001]	54..523 603..1072	469/470 (99%) 469/470 (99%)	0.0

In a BLAST search of public sequence databases, the NOV52a protein was found to have homology to the proteins shown in the BLASTP data in Table 52E.

Table 52E. Public BLASTP Results for NOV52a				
Protein Accession Number	Protein/Organism/Length	NOV52a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
CAC37778	Sequence 3 from Patent WO0123561 - Homo sapiens (Human), 523 aa.	1..523 1..523	523/523 (100%) 523/523 (100%)	0.0
Q9H324	ADAMTS-10 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase with thrombospondin motifs 10) (ADAM-TS 10) (ADAM-TS10) - Homo sapiens (Human), 1077 aa (fragment).	54..523 608..1077	469/470 (99%) 469/470 (99%)	0.0
P58459	ADAMTS-10 (EC 3.4.24.-) (A disintegrin and metalloproteinase with thrombospondin motifs 10) (ADAM-TS 10) (ADAM-TS10) - Mus musculus (Mouse), 450 aa (fragment).	75..522 1..449	416/449 (92%) 424/449 (93%)	0.0
CAC37777	Sequence 1 from Patent WO0123561 - Homo sapiens (Human), 634 aa (fragment).	54..465 223..634	412/412 (100%) 412/412 (100%)	0.0
CAD20434	Sequence 8 from Patent WO0188156 - Homo sapiens (Human), 1044 aa (fragment).	54..464 634..1044	411/411 (100%) 411/411 (100%)	0.0

PFam analysis predicts that the NOV52a protein contains the domains shown in the Table 52F.

Table 52F. Domain Analysis of NOV52a
--------------------------------------

Pfam Domain	NOV52a Match Region	Identities/ Similarities for the Matched Region	Expect Value
tsp_1	249..304	11/60 (18%) 38/60 (63%)	0.043
tsp_1	308..364	14/64 (22%) 38/64 (59%)	0.1
tsp_1	366..422	16/58 (28%) 34/58 (59%)	0.4
tsp_1	427..477	17/56 (30%) 32/56 (57%)	0.073

**Example 53.**

The NOV53 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 53A.

Table 53A. NOV53 Sequence Analysis			
	SEQ ID NO: 399	2245 bp	
NOV53a, CG56155-01 DNA Sequence	AGAACAGCTTGAAGACCGTTTCATTTTTAAGTGACAAGAGACTCACCTCCAAGAAGCAA TTGTGTTTTTCAAGATGATTTTATTCAAGCAAGCAACTTATTTTCATTTCTTGTGTTGCT ACAGTTTCTGTGGATGCTGACTCAACTCTATGAAAACGCCTTCTTCAGAGGTGGGG ATGTAGCTTCCATGTACACCCCAAATGCCCAATACTGCCAGATGAGGTGCACATTCCA CCCAAGGTGTTTGTATTTCAGTTTTCTTCCAGCAAGTTCAATCAATGACATGGAGAAA AGGTTTGGTTGCTTCTTGAAGATAGTGTACAGGAACCTGCCAAAAGTACATCGAA CAGGTGCAGTTTCTGGACATTCTTGAAGCAATGTGGTCATCAATAAGTGCTTGGCA TCGAGACATTTATAAAGGAGTTGATATGAGAGGAGTCAATTTTAAATGTGTCTAAGGTT AGCAGTGTGGAAGATGCCAAAAAGGTGCACCAATAACATTTCGCTGCCAGTTTTTTT CATATGCCACGCAAAACATTTACAAGGCAGAGTACCGGAACAATTGCCTATTAAAGTA CAGTCCCGGAGGAACACCTACCGCTATAAAGGTGCTGAGTAACGTGGAATCTGGATTCT TCACCTGAAGCCCTGTGCCCTTTTCAGAAATTGGTTGCCCATGAACATCTTCCAGCATC TTGCGTTCTCAGATGTGGATGTTGCCAGGGTTCTCACTCCAGATGCTTTTGTGTGTCG GACCATCTGCACCTATCACCCCAACTGCCTCTTCTTTACATTCTATACAAATGTATGG AAAATCGAGTCAAAAGAAATGTTTGTCTTCTTAAACATCTGAAAGTGGCACACCAA GTTCTCTACTCTCAAGAAAACACCATATCTGGATATAGCCTTTTAACTGCAAAAAG AACTTTACCTGAACCTGCCATTCTAAAATTACCCGGGAGTTGACTTTGGAGGAGAA GAATTGAATGTGACTTTTGTAAAGGAGTGAATGTTTGCCAAGAGACTTGCACAAAGA TGATTCGCTGTCAGTTTTTCACTTATTCTTTACTCCAGAAGACTGTAAGGAAGAGAA GTGTAAGTGTCTTAAAGATTATCTATGGATGGTTCTCCAACAGGATTGCGTATGGG ACACAAGGGAGCTCTGGTTACTCTTTGAGATTGTGTAACACTGGGGACAACCTGTCT GCACAACAAAACAAGCACACGCATTGTTGGAGGAACAACTCTTCTGGGGAGAGTG GCCCTGGCAGGTGAGCCTGCAGGTGAAGCTGACAGCTCAGAGGCACCTGTGTGGAGGG TCACTCATAGGACACCAAGTGGGTCCTCACTGCTGCCCACTGCTTTGATGGGCTTCCCC TGCAGGATGTTTGGCGCATCTATAGTGGCATTTAAATCTGTGAGATTACAAAAGA TACACCTTTCTCAAAATAAAAGAGATTATTATTCACCAAACTATAAAGTCTCAGAA GGAATCATGATATCGCCTTGATAAACTCCAGGCTCCTTTGAATTACACTGAATTCC AAAACCAATATGCCTACCTTCAAAGGTGACACAAGCACAATTTATACCAACTGTTG GGTAACCGGATGGGGCTTCTCGAAGGAGAAAGGTGAAATCCAAAATATTCTACAAAAG GTAAATATTCTTTGGTAACAAATGAAGATGCCAGAAAAGATATCAAGATTATAAAA TAACCAACGGATGGTCTGTGCTGGCTATAAAGAAGGGGAAAAGATGCTTGTAAAGGG AGATTCAAGGTGGTCCCTTAGTTTGCAACACAACGGAATGTGGCGTTTGGTGGGCATC ACAAGCTGGGGTGAAGGCTGTGCCCGAGGGAGCAACCTGGTGTCTACACCAAGTCG		

	CTGAGTACATGGACTGGATTTTAGAGAAAACACAGAGCAGTGATGGAAAAGCTCAGAT GCAGTCACCAGCATGAGAAGCAGTCCAGAGTCTAGGCAATTTTACAACTGAGTTCA AGTCAAATTTCTGAGCCTGGGGGGTCTCATCTGCAAAGCATGGAGAGTGGCATCTTCT TTGCATCCTAAGGACGAAAGACACAGTGCAGTCTGAGAGCTCTGAGGACAATGTCTGCT GAAGCCCGCTTTTACAGCAGCCGTAACCAGGGGCTGACAATGCGAGGTGCGAACTGAGA TCTCCATGACTGTGTGTTGTGAAATAAAATGGTGAAAGATC		
	ORF Start: ATG at 72		ORF Stop: TGA at 1986
	SEQ ID NO: 400	638 aa	MW at 71369.0kD
NOV53a, CG56155-01 Protein Sequence	MILFKQATYFISLFATVSCGCLTQLYENAFFRGGDVASMYTPNAQYQMRCTFHPRL LFSFLPASSINDMEKRFGCFLKDSVTGTLPKVHRTGAVSGHSLKQCGHQISACHRDIY KGVDMRGVNFNVSKVSSVEECQKRCTNNIRCOFFSYATQTFHKAERYNNCLLKYSPPG TPTAIKVLNSVESGFSCLKPALSEIGCHMNIQHLAFSDVDVARVLTDAFVCRITCT YHPNCLFFTFYTNVWKIESQRNVCLLKTSESGTPSSSTPQENTISGYSLLTCKRTLPE PCHSKIYPGVDFGGEELNVTFVKGVNVQCETCTKMIRCQFFYSLLPEDCKEEKCKCF LRLSMDGSPTRIAYGTTQSSGYSRLCNTGDNVCTTKTSTRIVGGTNSSWGEWPWQV SLQVKLTAQRHLCCGSLIGHQWVLTAAHCFDGLPLQDVWRIYSGIILNSDITKDTFPS QIKEIIHQNYKVSEGNHDIALIKLQAPLNYTEFQKPICLPSKGDSTIYTNCWVTGW GFSKEKEIQNILQKVNIPLVTNBECQKRYQDYKITQRMVCAGYKEGGKDACGDSGG PLVCKHNGMWRLVGITSWGEGCARREQPGVYTKVAEYMDWILEKTQSSDGKAQMQSPA		
	SEQ ID NO: 401	2038 bp	
NOV53b, CG56155-02 DNA Sequence	GTTTTCAGAAATGATTTTATTCAAGCAAGCAACTATTTCATTTCTTGTGTTGTACAG TTTCTGTGGATGTCTGACTCACTCTATGAAAACGCCTTCTTCAGAGGTGGGGATGT AGCTTCCATGTACACCCCAAATGCCCAATACTGCCAGATGAGGTGCACATTTCCACCA AGGTGTTTGCTATTCAAGTTTCTTCCAGCAAGTCAATCAATGACATGGAGAAAAGGT TTGGTTGCTTCTTGAAAGATAGTGTACAGGAACCTGCCAAAAGTACATCGAACAGG TGCAGTTTCTGGACATTCCTTGAAGCAATGTGGTCATCAAATAAGTGCTTGCCATCGA GACATTTATAAAGGAGTTGATATGAGAGGAGTCAATTTTAATGTGTCTAAGGTAGCA GTGTTGAAGAATGCCAAAAAGGTGCACCAATAACATTGCTGCCAGTTTTTTTCATA TGCCACGCAAAACATTTCAAGGCAGAGTACCGGAACAATTGCCATTAAAGTACAGT CCCGGAGGAACACCTACCGCTATAAAGGTGCTGAGTAACGTGGAATCTGGATTCTCAC TGAAGCCCTGTGCCCTTTTCAAGAAATGGTTGCCACATGAACATCTTCAGCATCTTGC GTTCTCAGATGTGGATGTTGCCAGGTTTCTCACTCCAGATGCTTTGTGTGTCGGACC ATCTGCACCTATCACCCCACTGCCCTTCTTTTACATTCTATACAAATGTATGGAAAA TCGAGTCACAAAGAAATGTTGTCTTCTTAAACATCTGAAAGTGGCACACCAAGTTC CTCTACTCCTCAAGAAAACACCATATCTGGATATAGCCTTTTAACTGCAAAAGAACT TTACCTGAACCTGCCATTCTAAAATTACCCGGGAGTTGACTTTGGAGGAGAAGAAT TGAATGTGACTTTTGTAAAGGAGTGAATGTTTGCCAAGAGACTTGCACAAAGATGAT TCGCTGTCAGTTTCTTACTTTTACTCCAGAAGACTGTAAGGAAGAGAAGTGT AAGTGTCTTAAAGATTATCTATGGATGTTCTCCAAGTGGATGCGTATGGGACAC AAGGGAGCTCTGGTTACTCTTTGAGATTGTGTAACACTGGGGACAACGCTGTCTGCAC AACAAAAACAAGCACACGCATTGTTGGAGGAACAACTCTTCTTGGGGAGAGTGGCCC TGGCAGGTGAGCCTGCAGGTGAAGCTGACAGCTCAGAGGCACCTGTGTGGAGGGTCAC TCATAGGACACCAGTGGGTCTCACTGCTGCCACTGCTTTGATGGGCTTCCCCTGCA GGATGTTTGGCGCATCTATAGTGGCATTTTAAATCTGTGACATTTACAAAAGATACA CCTTCTCACAAATAAAGAGATTATTATTCACCAAACTATAAAGTCTCAGAAGGGA ATCATGATATCGCTTGATAAACTCCAGGCTCCTTTGAATTACACTGAATTCAAAA ACCAATATGCCCTACCTTCCAAAGGTGACACAAGCACAATTTATACCAACTGTTGGGTA ACCGGATGGGGCTTCTCGAAGGAGAAAGGTGAAATCCAAAATATTCTACAAAAGGTAA ATATTCTTTGGTAACAAATGAAGAATGCCAGAAAAGATATCAAGATTATAAAATAAC CCAACGGATGGTCTGTGCTGGCTATAAAGAAGGGGAAAAGATGCTTGTAAAGGAGAT TCAGGTGGTCCCTTAGTTTGCAAAACACACGAATGTGGCGTTTGGTGGGCATCACCA GCTGGGGTGAAGGCTGTGCCCGCAGGGAGCAACCTGGTGTCTACACCAAGTGCCTGA GTACATGGACTGGATTTTAGAGAAAACACAGAGCAGTGATGGAAAAGCTCAGATGCAG TCACCAGCATGAGAAGCAGTCCAGAGTCTAGGCAATTTTACAACTGAGTTCAAGTC AAATCTGAGCCTGGGGGGTCTCATCTGCAAAGCATGAAGAGTGGCATCTTCTTTGCG ATCCTAAG		
	ORF Start: ATG at 10		ORF Stop: TGA at 1924

	SEQ ID NO: 402	638 aa	MW. at 71401.1kD
NOV53b, CG56155-02 Protein Sequence	MILFKQATYFISLFATVSCGCLTQLYENAFFRGGDVASMYTPNAQYQCMRCTFHPRLCLFSFLPASSINDMEKRFGCFLKDSVTGTLPKVHRTGAVSGHSLKQCGHQISACHRDIYKGVDMRGVNFNVSKVSSVEECQKRCTNNIRCQFFSYATQTFHKAERYNNCLLKSPGGTPTAIKVLNSNVESGFSLKPCALSEIGCHMNI FQHLAFSDVDVARFLTPDAFVCRITICTYHPNCLFFTFYTNVWKIESQRNVCLLKTSESGTPSSSTPQENTISGYSLLTCKRTLPEPCHSKIYPGVDFGGEELNVTFVKGVNVCQETCTKMIRCQFFTYSLLPEDCKEEKCKCFLRLSMDGSPTRIA YGTQGSSGYSLRLCNTGDNAVCTTKTSTRIVGGTNSSWGEWPWQVSLQVKLTAQRHLCCGSLIGHQWVLTAAHCFDGLPLQDVWRIYSGILNLSDITKDTFFS QIKEII IHQNYKVSEGNHDIALIKLQAPLNYTEFQKPICLPSKGDSTIYTNCWVTGWGFSKEKGEIQNILQKVNIPLVTNEECQKRYQDYKITQRMVCAGYKEGGKDACGDSGGPLVCKHNGMWRLVGITSWGEGCARREQPGVYTKVAEYMDWILEKTQSSDGKAQMQSPA		
	SEQ ID NO: 403	1869 bp	
NOV53c, CG56155-03 DNA Sequence	GGATCCGGATGTCTGACTCAACTCTATGAAAACGCCTTCTTCAGAGGTGGGGATGTAGCTTCCATGTACACCCCAATGCCAATACTGCCAGATGAGGTGCACATTCCACCCAAGGTGTTTGCTATTTCAGTTTCTTCCAGCAAGTTCAATCAATGACATGGAGAAAAGGTTTGGTTGCTTCTTGAAAGATAGTGTACAGGAACCTGCCAAAAGTACATCGAACAGGTGCAGTTTCTGGACATTCTCTGAAGCAATGTGGTCATCAAATAAGTGCTTGCCATCGAGACATTTATAAAGGAGTTGATATGAGAGGAGTCAATTTAATGTGTCTAAGGTTAGCAGTGTGAAGAATGCCAAAAAAGGTGCACCAAGTAACATTGCTGCCAGTTTTCATATGCCACGCAAAACATTTCAAGGCAGAGTACCGGAACAATGCCTATTAAAGTACAGTCCCGGAGGAACACCTACCGCTATAAAGGTGCTGAGTAACGTGAATCTGGATTCTCACTGAAGCCCTGTGCCCTTTCAGAAATTGGTTGCCACATGAACATCTTCAGCATCTTGCGTTCTCAGATGTGGATGTTGCCAGGTTTCTCACTCCAGATGCTTTGTGTGTGCGACCATCTGCACCTATCACCCCAACTGCCTCTTCTTTACATTCTATACAAATGTATGGAAAATCAGATCACAAAGAAATGTTTGTCTTCTTAAAACATCTGAAAGTGGCACACCAAGTTCCTCTACTCCTCAAGAAAACACCATACTGGATATAGCCTTTTAACTGCAAAAAGAACTTTACCTGAACCCCTGCCATTCTAAAATTTACCCGGGAGTTGACTTTGGAGGAGAAGAATTGAATGTGACTTTTGTAAAGGAGTGAATGTTTGCCAAGAGACTGCACAAAGATGATTGCTGTGACTTTTTCATTATTCTTTACTCCCAGAAGACTGTAAAGGAAGAGAAGTGTAAGTGTCTTAAAGATTATCTATGGATGGTTCTCCAAGTAGGATTGCGTATGGGACACAAAGGAGCTCTGGTTACTCTTTGAGATTGTGTAACACTGGGGACAACGCTGTCTGCACAACAAAAACAAGCACACGCATTGTTGAGGAACAACTCTTCTTGGGGAGAGTGGCCCTGCGAGGTGAGCCTGCAGGTGAAGCTGACAGCTCAGAGGCACCTGTGTGGAGGGTCACTCATAGGACACCAAGTGGTCTCTGCTGCCACTGCTTTGATGGGCTTCCCTGCAGGATGTTTGGCGCATCTATAGTGGCATTTTAAATCTGTGACACATTACAAAAGATACACCTTTCTCACAAATAAAGAGATTATTATTCACCAAACTATAAAGTCTCAGAAGGGAATCATGATATCGCCTTGATAAACTCCAGGCTCCTTGAATTACACTGAATTCAAAAAACCAATATGCCTACCTTCAAAGGTGACACAAGCACAATTTATACCAACTGTTGGGTAACCGGATGGGGCTTCTCGAAGGAGAAAGGTGAAATCCAAAATATTCTACAAAAGGTAATATTCTTTGGTAACAAATGAAGAATGCCAGAAAAGATATCAAGATTATAAATAACCAAACGATGGTCTGTGCTGGCTATAAAGAAGGGGAAAAGATGCTTGTAAGGGAGATTCAGGTGGTCCCTTAGTTGCAAACACAATGGAATGTGGCGTTTGGTGGGCATCACCAGCTGGGGTGAAGGCTGTGCCCGCAGGGAGCAACCTGGTGTCTACACCAAGTCGCTGAGTACATGGACTGGATTTTAGAGAAAACAGAGCAGTGATGGAAAAGCTCAGATGCAGTCAACGCACTCGAG		
	ORF Start: at 7		ORF Stop: at 1864
	SEQ ID NO: 404	619 aa	MW at 69208.4kD
NOV53c, CG56155-03 Protein Sequence	GCLTQLYENAFFRGGDVASMYTPNAQYQCMRCTFHPRLCLFSFLPASSINDMEKRFGCFLKDSVTGTLPKVHRTGAVSGHSLKQCGHQISACHRDIYKGVDMRGVNFNVSKVSSVEECQKRCTSNIRCQFFSYATQTFHKAERYNNCLLKSPGGTPTAIKVLNSNVESGFSLKPCALSEIGCHMNI FQHLAFSDVDVARFLTPDAFVCRITICTYHPNCLFFTFYTNVWKIESQRNVCLLKTSESGTPSSSTPQENTISGYSLLTCKRTLPEPCHSKIYPGVDFGGEELNVTFVKGVNVCQETCTKMIRCQFFTYSLLPEDCKEEKCKCFLRLSMDGSPTRIA YGTQGSSGYSLRLCNTGDNAVCTTKTSTRIVGGTNSSWGEWPWQVSLQVKLTAQRHLCCGSLIGHQWVLTAAHCFDGLPLQDVWRIYSGILNLSDITKDTFFS QIKEII IHQNYKVSEGNHDIALIKLQAPLNYTEFQKPICLPSKGDSTIYTNCWVTGWGFSKEKGEIQNILQKVNIPLVTNEECQKRYQDYKITQRMVCAGYKEGGKDACGDSGGPLVCKHNGMWRLVGITSWG		

EGCARREQPGVYTKVAEYMDWILEKTQSSDGKAQMSPA
--

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 53B.

Table 53B. Comparison of NOV53a against NOV53b and NOV53c.		
Protein Sequence	NOV53a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV53b	1..638	636/638 (99%)
	1..638	637/638 (99%)
NOV53c	20..638	616/619 (99%)
	1..619	618/619 (99%)

Further analysis of the NOV53a protein yielded the following properties shown in Table 53C.

Table 53C. Protein Sequence Properties NOV53a	
PSort analysis:	0.3700 probability located in outside; 0.1900 probability located in lysosome (lumen); 0.1000 probability located in endoplasmic reticulum (membrane); 0.1000 probability located in endoplasmic reticulum (lumen)
SignalP analysis:	Cleavage site between residues 20 and 21

- 5 A search of the NOV53a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 53D.

Table 53D. Geneseq Results for NOV53a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV53a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAU68928	Human protease domain of kallikrein I - Homo sapiens, 158 aa. [US6294663-B1, 25-SEP-2001]	427..584 1..158	158/158 (100%) 158/158 (100%)	1e-92
AAU82755	Amino acid sequence of novel human protease #54 - Homo sapiens, 802 aa. [WO200200860-A2, 03-JAN-2002]	319..621 513..797	115/306 (37%) 172/306 (55%)	9e-57
AAB24052	Human PRO618 protein sequence. SEQ ID NO:24 - Homo sapiens, 802 aa. [WO200053754-A1, 14-SEP-2000]	319..621 513..797	115/306 (37%) 172/306 (55%)	9e-57

AAB44266	Human PRO618 (UNQ354) protein sequence SEQ ID NO:169 - Homo sapiens, 802 aa. [WO200053756-A2, 14-SEP-2000]	319..621 513..797	115/306 (37%) 172/306 (55%)	9e-57
AAY41710	Human PRO618 protein sequence - Homo sapiens, 802 aa. [WO9946281-A2, 16-SEP-1999]	319..621 513..797	115/306 (37%) 172/306 (55%)	9e-57

In a BLAST search of public sequence databases, the NOV53a protein was found to have homology to the proteins shown in the BLASTP data in Table 53E.

Table 53E. Public BLASTP Results for NOV53a				
Protein Accession Number	Protein/Organism/Length	NOV53a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
P03952	Plasma kallikrein precursor (EC 3.4.21.34) (Plasma prekallikrein) (Kininogenin) (Fletcher factor) - Homo sapiens (Human), 638 aa.	1..638 1..638	638/638 (100%) 638/638 (100%)	0.0
O97506	Kallikrein - Sus scrofa (Pig), 643 aa.	1..635 9..643	505/635 (79%) 569/635 (89%)	0.0
Q8R0P5	Kallikrein B, plasma 1 - Mus musculus (Mouse), 638 aa.	1..638 1..638	487/638 (76%) 555/638 (86%)	0.0
P26262	Plasma kallikrein precursor (EC 3.4.21.34) (Plasma prekallikrein) (Kininogenin) (Fletcher factor) - Mus musculus (Mouse), 638 aa.	1..638 1..638	486/638 (76%) 554/638 (86%)	0.0
P14272	Plasma kallikrein precursor (EC 3.4.21.34) (Plasma prekallikrein) (Kininogenin) (Fletcher factor) - Rattus norvegicus (Rat), 638 aa.	1..638 1..638	478/638 (74%) 550/638 (85%)	0.0

PFam analysis predicts that the NOV53a protein contains the domains shown in the Table 53F.

Table 53F. Domain Analysis of NOV53a			
Pfam Domain	NOV53a Match Region	Identities/ Similarities for the Matched Region	Expect Value
PAN	21..104	19/112 (17%) 66/112 (59%)	6.8e-14

PAN	111..194	24/111 (22%) 67/111 (60%)	5.4e-15
PAN	201..284	21/111 (19%) 63/111 (57%)	1.3e-10
PAN	292..375	23/111 (21%) 64/111 (58%)	2.3e-09
trypsin	391..621	113/262 (43%) 196/262 (75%)	4.8e-100

**Example 54.**

The NOV54 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 54A.

Table 54A. NOV54 Sequence Analysis			
	SEQ ID NO: 405	1010 bp	
NOV54a, CG57191-01 DNA Sequence	CGTATTGCTCGGCCCGGGGAGTTTCGCCCCCTGCCCGGCTCCGCGGCGCGGAGGATGG CGTGGAACGGCTGGGCGCGCTGGTGATGTTCCCTCTACAGATGATCTATCTGGTGGT GAAAGCAGCCGTCGACTGGTGCTGCCCGCCAAGCTGCGGGACCTGTCGCGGGAGAAC GTCCTCATCACCGGCGGCGGGAGAGGCATCGGGCGTCAGCTCGCCCGGAGTTCGCGG AGCGCGCGCCAGAAAGATTGTTCTCTGGGGCCGACTGAGAAATGCCTGAAGGAGAC GACGGAGGAGATCCGGCAGATGGGCACTGAGTGCCATTACTTCATCTGTGATGTGGGC AACCGGAGAGGTGTACAGACGGCCAAGCCGTCGCGGAGAAGGTGGGTGACATCA CCATCCTGGTGAACAATGCCCGCGTGGTCCATGGGAAGAGCCTAATGGACAGTGATGA TGATGCCCTCCTCAAGTCCCAACACATCAACACCCTGGGCCAGTTCCTGGACCACCAAG GCCTTCCTGCCGCTATGCTGGAGCTGCAGAATGGCCACATCGTGTGCCTCAACTCCG TGCTGGCACTGTCTGCCATCCCCGGTGCCATCGACTACTGCACATCCAAAGCGTCAGC CTTGCGCTTCATGGAGAGCCTGACCCTGGGGCTGCTGGACTGTCCGGGAGTCAGCGCC ACCACAGTGTGCGCTTCCACACCAGCACCAGAGATGTTCCAGGGCATGAGAGTCAGGT TTCCCAACCTCTTTCCCCCACTGAAGCCGGAGACGGTGGCCCGGAGGACAGTGGAAGC TGTGCAGCTCAACCAGGCCCTCCTCCTCCCATGGACAATGCATGCCCTCGTTATC TTGAAAAGCATACTTCCACAGGCTGCACTCGAGGAGATCCACAAATCTCAGGAACCT ACACCTGCATGAACACTTTCAAAGGGCGGACATAGAGACAGGATGAAGACATGCTTGA GGAGCCACGGAGTTTGGGGGCCAC		
	ORF Start: ATG at 55		ORF Stop: TAG at 961
	SEQ ID NO: 406	302 aa	MW at 33520.0kD
NOV54a, CG57191-01. Protein Sequence	MAWKRLGALVMFPLQMIYLVVKAAGVGLVLPKLRDLRSRENVLTGGGRGIGRQLAREF AERGARKIVLWGRTEKCLKETTEEIRQMGTECHYFICDVGNREEVYQTAKAVREKVG ITILVNNAAVVHGKSLMDSDDDALLKSKHINTLGQFWTTKAFLLPRMLELQNGHIVCLN SVLALSAIPGAIDYCTSKASAFAMESLTLGLLDCPGVSATTVLPFHTSTEMFQGMRV RFPNLFPLPKPETVARRTVEAVQLNQALLLPWTMHALVILKSLPQAALKEEIHKFSG TYTCMNTFKGRT		
	SEQ ID NO: 407	1330 bp	
NOV54b, CG57191-03 DNA Sequence	GGAGTTTCGCCCCCTGCCCGGCTCCGCGGCGCGGAGGATGGTGTGGAACGGCTGGGC GCGCTGGTGATGTTCCCTCTACAGATGATCTATCTGGTGGTGAAAGCAGCCGTCGGAC TGGTGCTGCCCGCCAAGCTGCGGGACCTGTCGCGGGAGAACGTCCTCATCACCGGCGG CGGGAGAGGCATCGGGCGTCAGCTCGCCCGGAGTTCGCGGAGCGCGGCCAGAAAG ATTGTTCTCTGGGGCCGACTGAGAAATGCCATTACTTCATCTGTGATGTGGGCAACC GGGAGGAGGTGTACAGACGGCCAAGGCCGTCCGGGAGAAGGTGGGTGGCATCACCAT CCTGGTGAGCAATGCCCGCGTGGTCCATGGGAAGAGCCTAATGGACAGTGATGATGAT GCCTTCCTCAAGTCCCAACACATCAACACCCTGGGCCAGTTCCTGGACCACCAAGGCCT		

	TCCTGCCGCGTATGCTGGAGCTGCAGAATGGCCACATCGTGTGCCTCAACTCCGTGCT GGCACTGTCTGCCATCCCCGGTGCCATCGACTACTGCACATCCAAAGCGTCAGCCTTC GCCTTCATGGAGAGCCTGACCCTGGGGCTGCTGGACTGTCCGGGAGTCAGCGCCACCA CAGTGTGCCCCCTCCACACCAGCACCCGAGATGTTCCAGGGCATGAGAGTCAGGTTTCC CAACCTCTTTCCCCACTGAAGCCGGAGACGGTGGCCCCGGAGGACAGTGGAAAGCTGTG CAGCTCAACCAGGCCCTCCTCCTCCTCCATGGACAATGCATGCCCTCGTTATCTTGA AAAGCATACTTCCACAGGCTGCACTCGAGGAGATCCACAAATTCTCAGGAACCTACAC CTGCATGAACACTTTCAAAGGGCGGACATAGAGACAGGATGAAGACATGCTTGAGGAG CCACGGAGTTTGGGGGCCACAGCACCTGGGCACACACCCGAGCACCTGTCCATTGGCA TGCTTCTGCTGGGTGAGCAGGACAGCTCCTGTCCCCAGCGAAGAATCCGGCTGCCCTT GGGCCAGTCCCAGGACCTTTGCACAGGACTGATGGGTATAACTGACCCCAACAGGGAG GCAGGAAAACAGCCAGAAGCCACCTTGACACTTTTGAACATTTCCAGTTCTGTAGAGT TTATTGTCAATTGCTTCTCAAGTCTAACCAGCCTCAGCAGTGTGCATAGACCATTTC AGGAGGGTCTGTCCCAGATGCTCTGCCTCCCGTTCCAAAACCCACTCATCCTCAGCT TGCACAAACTGGTTGAACGGCAGGAATGAAAAATAAGAGAGATGGCTTTTGTG		
	ORF Start: ATG at 38		ORF Stop: TAG at 899
	SEQ ID NO: 408	287 aa	MW at 31731.0kD
NOV54b, CG57191-03 Protein Sequence	MVWKRLGALVMFPLQMIYLVVKAAGVGLVLPKLRDLRSRENVITGGGRGIGRQLAREF AERGARKIVLWGRTEKCHYFICDVGNREEVYQTAKAVREKVGGITILVSNAAVVHGKS LMDSDDDAFLKSQHINTLGQFWTTKAFLEPRMLELQNGHIVCLNSVLALSAIPGAIDYC TSKASAFAMESLTLGLLDCPGVSATTVLPFHTSTEMFQGMVRVFPNLFPLKPEVTA RRTVEAVQLNQALLLPWTMHALVILKSILPQAALIEIHKFSGYTCMNTFKGRT		
	SEQ ID NO: 409		992 bp
NOV54c, CG57191-02 DNA Sequence	GGAGTTTCGCCCCCTGCCCGGCTCCGCGGCGCGGAGGATGGTGTGGAACGGCTGGGC GCGCTGGTGATGTTCCCTCTACAGATGATCTATCTGGTGGTGAAAGCAGCCGTCGGAC TGGTGCTGCCCCCAAGCTGCGGGACCTGTGCGGGGAGAACGTCCTCATCACCGGCGG CGGGAGAGGCATCGGGCGTCAGCTCGCCCGCGAGTTCGCGGAGCGCGGCCGAGAAAG ATTGTTCTCTGGGGCCGACTGAGAAATGCCTGAAGGAGACGACAGAGGGGATCCGGC AGATGGGCACTGAGTGCCACTACTTCATCTGTGATGTGGGCAACCGGGAGGAGGTGTA CCAGACGGCCAAGGCCGTCCGGGAGAAGGTGGGTGACATCACCATCCTGGTGAACAAT GCCGCGTGGTCCATGGGAAGAGCCTAATGGACAGTGAATGATGATGCCCTCCTCAAGT CCCAACACATCAACACCCTGGGCCAGTTCTGGACCACCAAGGCCTCCTGCCGCGTAT GCTGGAGCTGCAGAATGGCCACATCGTGTGCCTCAACTCCGTGCTGGCACTGTCTGCC ATCCCCGCTGCCATCGACTACTGCACATCCAAAGCGTCAGCCTTCGCCTTCATGGAGA GCCTGACCCTGGGGCTGCTGGACTGTCCGGGAGTCAGCGCCACCAAGTGTGCTGCCCTT CCACACCAGCACCGAGATGTTCCAGGGCATGAGAGTCAGGTTTCCCAACCTCTTTCCC CCTGAAAGCCGGAGACGGTGGCCCGGAGGACAGTGAAGCTGTGCAGCTCAACCAGG CCCTCCTCCTCCTCCCATGGACAATGCATGCCCTCGTTATCTTGAAGCATACTTCC ACAGGCTGCACTCGAGGAGATCCACAAATTCTCAGGAACCTACACCTGCATGAACACT TTCAAAGGGCGGACATAGAGACAGGATGAAGACATGCTTGAGGAGCCACGGAGTTTGG GGGCCA		
	ORF Start: ATG at 38		ORF Stop: TAG at 944
	SEQ ID NO: 410	302 aa	MW at 33476.0kD
NOV54c, CG57191-02 Protein Sequence	MVWKRLGALVMFPLQMIYLVVKAAGVGLVLPKLRDLRSRENVITGGGRGIGRQLAREF AERGARKIVLWGRTEKCLKETTEGIRQMGTECHYFICDVGNREEVYQTAKAVREKVG DITILVNNAVVHGKSLMDSDDDAFLKSQHINTLGQFWTTKAFLEPRMLELQNGHIVCLN SVLALSAIPGAIDYCTSKASAFAMESLTLGLLDCPGVSATTVLPFHTSTEMFQGMVR RFPNLFPLKPEVARTVEAVQLNQALLLPWTMHALVILKSILPQAALIEIHKFSG TYTCMNTFKGRT		

5 Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 54B.

Table 54B. Comparison of NOV54a against NOV54b and NOV54c.

Protein Sequence	NOV54a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV54b	1..302	282/302 (93%)
	1..287	283/302 (93%)
NOV54c	1..302	300/302 (99%)
	1..302	300/302 (99%)

Further analysis of the NOV54a protein yielded the following properties shown in Table 54C.

Table 54C. Protein Sequence Properties NOV54a	
PSort analysis:	0.6850 probability located in endoplasmic reticulum (membrane); 0.6400 probability located in plasma membrane; 0.4600 probability located in Golgi body; 0.1000 probability located in endoplasmic reticulum (lumen)
SignalP analysis:	Cleavage site between residues 24 and 25

A search of the NOV54a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 54D.

5

Table 54D. Geneseq Results for NOV54a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV54a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAY92510	Human OXRE-7 - Homo sapiens, 302 aa. [WO200020604-A2, 13-APR-2000]	1..302 1..302	301/302 (99%) 301/302 (99%)	e-173
AAW89191	Bone morphogenetic protein upregulated gene (29A) product - Mus sp, 202 aa. [EP890639-A2, 13-JAN-1999]	1..195 1..196	177/196 (90%) 184/196 (93%)	2e-97
AAO05702	Human polypeptide SEQ ID NO. 19594 - Homo sapiens, 138 aa. [WO200164835-A2, 07-SEP-2001]	144..281 1..138	137/138 (99%) 137/138 (99%)	3e-74
AAY97999	Human SCAD family molecule HSFM-1, SEQ ID NO:1 - Homo sapiens, 309 aa. [US6057140-A, 02-MAY-2000]	9..298 11..302	105/293 (35%) 167/293 (56%)	2e-47
ABB72322	Rat protein isolated from skin cells SEQ ID NO: 646 - Rattus sp. 298	6..301 5..298	99/299 (33%) 170/299 (56%)	3e-46

	aa. [WO200190357-A1, 29-NOV-2001]			
--	-----------------------------------	--	--	--

In a BLAST search of public sequence databases, the NOV54a protein was found to have homology to the proteins shown in the BLASTP data in Table 54E.

Table 54E. Public BLASTP Results for NOV54a				
Protein Accession Number	Protein/Organism/Length	NOV54a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
O75911	Retinal short-chain dehydrogenase/reductase RETSDR1 (EC 1.-.-) - Homo sapiens (Human), 302 aa.	1..302 1..302	302/302 (100%) 302/302 (100%)	e-173
Q9BUC8	Short-chain dehydrogenase/reductase 1 - Homo sapiens (Human), 302 aa.	1..302 1..302	301/302 (99%) 301/302 (99%)	e-173
O77769	Retinal short-chain dehydrogenase/reductase RETSDR1 (EC 1.-.-) - Bos taurus (Bovine), 302 aa.	1..302 1..302	297/302 (98%) 300/302 (98%)	e-171
Q91WR0	Retinal short-chain dehydrogenase/reductase 1 - Mus musculus (Mouse), 302 aa.	1..302 1..302	286/302 (94%) 294/302 (96%)	e-165
Q91XC3	Similar to retinal short-chain dehydrogenase/reductase 1 - Mus musculus (Mouse), 302 aa.	1..302 1..302	285/302 (94%) 293/302 (96%)	e-165

PFam analysis predicts that the NOV54a protein contains the domains shown in the Table 54F.

Table 54F. Domain Analysis of NOV54a			
Pfam Domain	NOV54a Match Region	Identities/ Similarities for the Matched Region	Expect Value
adh_short	37..292	67/284 (24%) 171/284 (60%)	1.1e-25

## 5 Example 55.

The NOV55 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 55A.

Table 55A. NOV55 Sequence Analysis			
	SEQ ID NO: 411	1192 bp	
NOV55a, CG59595-01 DNA Sequence	CGGCGACTGACCGTGGTTCGTGGGCGGACGGCGGCTTGCAGCGTGGAGGAGCTGGGGTCT GCTGTGGGTTCGCGAAGCAGAGCCCGGGACGTGCGCGCTTGGTGACGATCCTGAAGGG GAGCTCCGAGGGGCGCGGTTCGCCAGGGCTGCTGCGGCCATTCCCGAGCCCGGCGCG GGGCCCGGAGATACTGGTTTAGGCCGTCCAGGGCTCCGGGCGCACCCGGTGGCCGC TGCTGCAGCGGAGGGAGCGCGGCGCGCGGGGCTCGGAGACAGCGTTTCTCCCGGAA GTCTTCTCCGGGCGAGAGTGGGAAGTGGGAGCCGGAGCGGCAGCTGGCAGCGTTCTC TCCGCAGGTCCGCACCATGCGCCCTGCAGCCCTGCGCGGGGCTGCTGGGCTGCCTC TGCCTGGCGTTGCTTTGCTTGGGCGGTGCGGACAAAGCGCTGCTGACAACCATGAGT GGAAAAAATAATTATGGTTCAGCACTGGCCTGAGACAGTATGCGAGAAAAATCAAAA CGACTGTAGAGACCCCTCCGATTACTGGACAATACATGGACTATGGCCCGATAAAAGT GAAGGATGTAATAGATCGTGGCCCTTCAATTTAGAAGAGATTAAGGATCTTTTGCCAG AAATGAGGGCATACTGGCCTGACGTAATTCACGTTTCCCAATCGCAGCCGCTTCTG GAAGCATGAGTGGGAAAAGCATGGGACCTGCGCGCCCGAGGTGGATGCGCTCAACTCC CAGAAGAAGTACTTTGGCAGAAGCCTGGAAGTCTACAGGGAGCTGGACCTCAACAGTG TGCTTCTAAAATTGGGGATAAAACCATCCATCAATTACTACCAAGTTGCAGATTTAA AGATGCCCTTGCCAGAGTATATGGAGTGATACCCAAAATCCAGTGCCCTTCCACCAAGC CAGGATGAGGAAGTACAGACAATTGGTCAGATAGAAGTGTGCCTCACTAAGCAAGACC AGCAGCTGCAAACTGCACCGAGCCGGGGAGCAGCGCTCCCAAGCAGGAAGTCTG CTTGGCAAATGGGGCCGCGAGAGCCGGGGTCTGAGAGTCTGTGAAGATGGCCAGTCT TTCTATCCCCACCTAAAAGACCAAGCATTGATGCCCAAGTTTGGAAATATTCTGT TTTAAAAGCAAGAGAAATTCACAACTGCAG		
	ORF Start: ATG at 365		ORF Stop: TGA at 1133
	SEQ ID NO: 412	256 aa	MW at 29480.5kD
NOV55a, CG59595-01 Protein Sequence	MRPAALRGALLGCLCLALLCLGGADKRLRDNHEWKKLIMVQHWPETVCEKIQNDCRDP PDYWTIHGLWPKSEGCNRSWPFNLEEIKDLLPEMRAYPDVIHSFPNRSRFWKHEWE KHGTCAAQVDALNSQKKYFGRSLELYRELDLNSVLLKLGKPSINYQVADF KDALAR VYGVIPKIQCLPPSQDEEVQTIQIIECLTKQDQQLQNCTEPGEQPSPKQEVWLANGA AESRGLRVCEGDPVFYPPPKKTKH		
	SEQ ID NO: 413	708 bp	
NOV55b, 169728691 DNA Sequence	GGATCCGACAAGCGCCTGCGTGACAACCATGAGTGGAAAAAATAATTATGGTTCAGC ACTGGCCTGAGACAGTATGCGAGAAAAATCAAAACGACTGTAGAGACCCCTCCGATTA CTGGACAATACATGGACTATGGCCCGATAAAAGTGAAGGATGTAATAGATCGTGGCCC TTCAATTTAGAAGAGATTAAGGATCTTTTGCCAGAAATGAGGGCATACTGGCCTGACG TAATTCACGTTTCCCAATCGCAGCCGCTTCTGGAAGCATGAGTGGGAAAAGCATGG GACCTGCGCCGCCAGGTGGATGCGCTCAACTCCAGAAGAAGTACTTTGGCAGAAGC CTGGAAGTCTACAGGGAGCTGGACCTCAACAGTGTGCTTCTAAAATTGGGGATAAAAC CATCCATCAATTACTACCAAGTTGCAGATTTTAAAGATGCCCTTGCCAGAGTATATGG AGTGATACCCAAAATCCAGTGCCTTCCACCAAGCCAGGATGAGGAAGTACAGACAATT GGTCAGATAGAAGTGTGCCTCACTAAGCAAGACCAGCAGCTGCAAACTGCACCGAGC CGGGGGAGCAGCCGTCCCCAAGCAGGAAGTCTGGCTGGCAAATGGGGCCGCCGAGAG CCGGGGTCTGAGAGTCTGTGAAGATGGCCAGTCTTCTATCCCCACCTAAAAGACC AAGCATCTCGAG		
	ORF Start: at 1		ORF Stop: end of sequence
	SEQ ID NO: 414	236 aa	MW at 27528.0kD
NOV55b, 169728691 Protein Sequence	GSDKRLRDNHEWKKLIMVQHWPETVCEKIQNDCRDPDYWTIHGLWPKSEGCNRSWP FNLEEIKDLLPEMRAYPDVIHSFPNRSRFWKHEWEKHGTCAAQVDALNSQKKYFGRS LELYRELDLNSVLLKLGKPSINYQVADF KDALARVYGVIPKIQCLPPSQDEEVQTI GQIELCLTKQDQQLQNCTEPGEQPSPKQEVWLANGA AESRGLRVCEGDPVFYPPPKKT KHLE		
	SEQ ID NO: 415	709 bp	
NOV55c, 160728707 DNA	GGATCCGACAAGCGCCTGCGTGACAACCATGAGTGGAAAAGACTAATTATGGTTCAGC ACTGGCCTGAGACAGTATGCGAGAAAAATCAAAACGACTGTAGAGACCCCTCCGATTA		

169728707 DNA Sequence	CTGGACAATACATGGACTATGGCCCGATAAAAGTGAAGGATGTAATAGATCGTGGCCC TTCAATTTAGAAGAGATTAAGGGTCTTTTGGCAGAAATGAGGGCATACTGGCCTGACG TAATTCACCTCGTTTCCCAATCGCAGCCGCTTCTGGAAGCATGAGTGGGAAAAGCATGG GACCGGCGCCGCCAGGTGGATGCGCTCAACTCCCAGAAGAAGTACTTTGGCAGAAGC CTGGAACCTCTACAGGGGGCTGGACCTCAACAGTGTGCTTCTAAAATTGGGGATAAAAC CATCCATCAATTACTACCAAGTTGCAGATTTTAAAGATGCCCTTGGCAGAGTATATGG AGTGATACCCAAATCCAGTGCCTTCCACCAAGCCAGGATGAGGAAGTACAGACAATT GGTCAGATAGAAGTGTGCCTCACTAAGCAAGACCAGCAGCTGCAAACTGCACCGAGC CGGGGGAGCAGCCGTCCCCCAAGCAGGAAGTCTGGCTGGCAAATGGGGCCGCCGAGAG CCGGGTCTGAGAGTCTGTGAAGATGGCCAGTCTTCTATCCCCACCTAAAAAGACC AAGCATCTCGAGA		
	ORF Start: at 1	ORF Stop: end of sequence	
	SEQ ID NO: 416	237 aa	MW at 27379.8kD
NOV55c, 169728707 Protein Sequence	GSDKRLRDNHEWKRLIMVQHPETVCEKIQNDICRDPDYWTIHGLWPKSEGCNRSWP FNLEEI KGLLP EMRAYWPDVIHSFPNRSR FWKHEWEKHGTGAAQVDALNSQKKYFGRS LELYRGLDLSVLLKLG IKPSINYQVADF KDALARVYGVIPKIQLPPSQDEEVQTI GQIELCLTKQDQQLQNCTEPGEQSPKQEVWLANGAAESRGLRVCEDEGPVFYPPPKKT KHLEX		
	SEQ ID NO: 417	708 bp	
NOV55d, 169728746 DNA Sequence	GGATCCGACAAGCGCCTGCGTGACAACCATGAGTGGAAAAAATAATTATGGTTCAGC ACTGGCCTGAGACAGTATGCGAGAAAATTCAAGACGACTGTAGAGACCCTCCGGATTA CTGGACAATACATGGACTATGGCCCGATAAAAGTGAAGGATGTAATAGATCGTGGCCC TTCAATTTAGAAGAGATTAAGGATCTTTTGGCAGAAATGAGGGCATACTGGCCTGACG TAATTCACCTCGTTTCCCAATCGCAGCCGCTTCTGGAAGCATGAGTGGGAAAAGCATGG GACCTGCGCCGCCAGGTGGATGCGCTCAACTCCCAGAAGAAGTACTTTGGCAGAAGC CTGGAACCTCTACAGGGAGCTGGACCTCAACAGTGTGCTTCTAAAATTGGGGATAAAAC CATCCATCAATTACTACCAAGTTGCGGATTTTAAAGATGCCCTTGGCAGAGTATATGG AGTGATACCCAAATCCAGTGCCTTCCACCAAGCCGGATGAGGAAGTACAGACAATT GGTCAGATAGAAGTGTGCCTCACTAAGCAAGACCAGCAGCTGCAAACTGCACCGAGC CGGGGGAGCAGCCGTCCCCCAAGCAGGAAGTCTGGCTGGCAAATGGGGCCGCCGAGAG CCGGGTCTGAGAGTCTGTGAAGATGGCCAGTCTTCTATCCCCACCTAAAAAGACC AAGCATCTCGAG		
	ORF Start: at 1	ORF Stop: end of sequence	
	SEQ ID NO: 418	236 aa	MW at 27557.0kD
NOV55d, 169728746 Protein Sequence	GSDKRLRDNHEWKRLIMVQHPETVCEKIQDDICRDPDYWTIHGLWPKSEGCNRSWP FNLEEI KDLLPEMAYWPDVIHSFPNRSR FWKHEWEKHGTCAAQVDALNSQKKYFGRS LELYRELDLSVLLKLG IKPSINYQVADF KDALARVYGVIPKIQLPPSRDEEVQTI GQIELCLTKQDQQLQNCTEPGEQSPKQEVWLANGAAESRGLRVCEDEGPVFYPPPKKT KHLE		
	SEQ ID NO: 419	708 bp	
NOV55e, CG59595-02 DNA Sequence	GGATCCGACAAGCGCCTGCGTGACAACCATGAGTGGAAAAAATAATTATGGTTCAGC ACTGGCCTGAGACAGTATGCGAGAAAATTCAAACGACTGTAGAGACCCTCCGGATTA CTGGACAATACATGGACTATGGCCCGATAAAAGTGAAGGATGTAATAGATCGTGGCCC TTCAATTTAGAAGAGATTAAGGATCTTTTGGCAGAAATGAGGGCATACTGGCCTGACG TAATTCACCTCGTTTCCCAATCGCAGCCGCTTCTGGAAGCATGAGTGGGAAAAGCATGG GACCTGCGCCGCCAGGTGGATGCGCTCAACTCCCAGAAGAAGTACTTTGGCAGAAGC CTGGAACCTCTACAGGGAGCTGGACCTCAACAGTGTGCTTCTAAAATTGGGGATAAAAC CATCCATCAATTACTACCAAGTTGCAGATTTTAAAGATGCCCTTGGCAGAGTATATGG AGTGATACCCAAATCCAGTGCCTTCCACCAAGCCAGGATGAGGAAGTACAGACAATT GGTCAGATAGAAGTGTGCCTCACTAAGCAAGACCAGCAGCTGCAAACTGCACCGAGC CGGGGGAGCAGCCGTCCCCCAAGCAGGAAGTCTGGCTGGCAAATGGGGCCGCCGAGAG CCGGGTCTGAGAGTCTGTGAAGATGGCCAGTCTTCTATCCCCACCTAAAAAGACC AAGCATCTCGAG		

	ORF Start: at 7		ORF Stop: at 703
	SEQ ID NO: 420	232 aa	MW at 27141.6kD
NOV55e, CG59595-02 Protein Sequence	DKRLRDNHEWKKLIMVQHWPETVCEKIQNDCRDPPDYWTIHGLWPKDSEGCNRSWPFN LEEIKDLLPEMRAYWPDVIHSFPNRSRFWKHEWEKHGTCAAQVDALNSQKKYFGRSLE LYRELDLNSVLLKLGIKPSINYQVADFKDALARVYGVI PKIQCLPPSQDEEVQTIGQ IELCLTKQDQQLQNCTEPGEQPSPKQEVWLANGAAESRGLRVCEGDGPVFYPPPKKTKH		
	SEQ ID NO: 421	923 bp	
NOV55f, CG59595-03 DNA Sequence	GAGACAGCGTTTCTCCCGAAGTCTCTCGGGCAGCAGGTGGGAAGTGGGAGCCGGA GCGGCAGCTGGCAGCGTTCTCTCCGAGGTGCGCACCATGCGCCCTGCAGCCCTGCGC GGGGCCCTGCTGGGCTGCCTCTGCCTGGCGTTGCTTTGCCTGGGCGGTGCGGACAAGC GCCTGCGTGACAACCATGAGTGGAAAACTAATTATGGTTACGACTGGCCTGAGAC AGTATGCGAGAAAATTCAAAACGACTGTAGAGACCTCCGGATTACTGGACAATACAT GGACTATGGCCCGATAAAAGTGAAGGATGTAATAGATCGTGGCCCTTCAATTTAGAAG AGATTAAGGATCTTTTGGCAGAAATGAGGGCATACTGGCCTGACGTAATTCACCTCGTT TCCAATCGCAGCCGCTTCTGGAAGCATGAGTGGGAAAAGCATGGGACCTGCGCCGCC CAGGTGGATGCGCTCACTCCAGAGAAGTACTTTGGCAGAAGCCTGGAACCTCTACA GGGAGCTGGACCTCAACAGTGTGCTTCTAAAATTTGGGGATAAAACCATCCATCAATTA CTACCAAGTTGCAGATTTTAAAGATGCCCTCGCCAGAGTATATGGAGTGATACCCAAA ATCCAGTGCCTTCCACCAAGCCAGGATGAGGAAGTACAGACAATTGGTCAGATAGAAC TGTGCCTCACTAAGCAAGACCAGCAGCTGCAAACTGCACCGAGCCGGGGGAGCAGCC GTCCCCAAGCAGGAAGTCTGGCTGGCAAATGGGGCCGCCGAGAGCCGGGGTCTGAGA GTCTGTGAAGATGGCCAGTCTTCTATCCCCCACTAAAAAGACCAAGCATTGATGCC CAAGTTTTGGAAATATTCTGTTTTAAAAAGCAAGAGAAATTCACAACTGCAG		
	ORF Start: ATG at 96		ORF Stop: TGA at 864
	SEQ ID NO: 422	256 aa	MW at 29480.5kD
NOV55f, CG59595-03 Protein Sequence	MRPAALRGALLGCLCLALLCLGGADKRLRDNHEWKKLIMVQHWPETVCEKIQNDCRDP PDYWTIHGLWPKDSEGCNRSWPFNLEEIKDLLPEMRAYWPDVIHSFPNRSRFWKHEWE KHGTCAAQVDALNSQKKYFGRSLELYRELDLNSVLLKLGIKPSINYQVADFKDALAR VYGVI PKIQCLPPSQDEEVQTIGQIELCLTKQDQQLQNCTEPGEQPSPKQEVWLANGA AESRGLRVCEGDGPVFYPPPKKTKH		
	SEQ ID NO: 423	709 bp	
NOV55g, CG59595-04 DNA Sequence	GGATCCGACAAGCGCCTGCGTGACAACCATGAGTGGAAAAGACTAATTATGGTTTCAGC ACTGGCCTGAGACAGTATGCGAGAAAATTCAAACGACTGTAGAGACCTCCGGATTA CTGGACAATACATGGACTATGGCCCGATAAAAAGTGAAGGATGTAATAGATCGTGGCCC TTCAATTTAGAAGAGATTAAGGGTCTTTTGGCAGAAATGAGGGCATACTGGCCTGACG TAATTCACCTCGTTTCCCAATCGCAGCCGCTTCTGGAAGCATGAGTGGGAAAAGCATGG GACCGGCGCCGCCCAGGTGGATGCGCTCAACTCCAGAGAAGTACTTTGGCAGAAGC CTGGAACCTACAGGGGCTGGACCTCAACAGTGTGCTTCTAAAATTTGGGGATAAAAC CATCCATCAATTACTACCAAGTTGCAGATTTTAAAGATGCCCTTGCCAGAGTATATGG AGTGATACCCAAAATCCAGTGCCTTCCACCAAGCCAGGATGAGGAAGTACAGACAATT GGTCAGATAGAACTGTGCCTCACTAAGCAAGACCAGCAGCTGCAAACTGCACCGAGC CGGGGGAGCAGCCGTCCCCAAGCAGGAAGTCTGGCTGGCAAATGGGGCCGCCGAGAG CCGGGGTCTGAGAGTCTGTGAAGATGGCCAGTCTTCTATCCCCACCTAAAAAGACC AAGCATCTCGAGA		
	ORF Start: at 7		ORF Stop: at 703
	SEQ ID NO: 424	232 aa	MW at 26993.4kD
NOV55g, CG59595-04 Protein Sequence	DKRLRDNHEWKRLIMVQHWPETVCEKIQNDCRDPPDYWTIHGLWPKDSEGCNRSWPFN LEEIKGLLPEMRAYWPDVIHSFPNRSRFWKHEWEKHGTGAAQVDALNSQKKYFGRSLE LYRGLDLNSVLLKLGIKPSINYQVADFKDALARVYGVI PKIQCLPPSQDEEVQTIGQ IELCLTKQDQQLQNCTEPGEQPSPKQEVWLANGAAESRGLRVCEGDGPVFYPPPKKTKH		
	SEQ ID NO: 425	708 bp	
NOV55h, CG59595-05 DNA Sequence	GGATCCGACAAGCGCCTGCGTGACAACCATGAGTGGAAAACTAATTATGGTTTCAGC ACTGGCCTGAGACAGTATGCGAGAAAATTCAAGACGACTGTAGAGACCTCCGGATTA CTGGACAATACATGGACTATGGCCCGATAAAAAGTGAAGGATGTAATAGATCGTGGCCC		

DNA Sequence	TTCAATTTAGAAGAGATTAAGGATCTTTTGCCAGAAATGAGGGCATACTGGCCTGACG TAATTCACCTCGTTTCCCAATCGAGCCGCTTCTGGAAGCATGAGTGGGAAAAGCATGG GACCTGCGCCGCCAGGTGGATGCGCTCAACTCCCAAGAAGTACTTTGGCAGAAGC CTGGAACCTCTACAGGGAGCTGGACCTCAACAGTGTGCTTCTAAAATTGGGGATAAAAC CATCCATCAATTACTACCAAGTTGCGGATTTTAAAGATGCCCTTGCCAGAGTATATGG AGTGATACCCAAAATCCAGTGCCTTCCACCAAGCCGGGATGAGGAAGTACAGACAATT GGTCAGATAGAACTGTGCCTCACTAAGCAAGACCAGCAGCTGCAAAACTGCACCGAGC CGGGGAGCAGCCGCTCCCCAAGCAGGAAGTCTGGCTGGCAAATGGGGCCCGCCGAGAG CCGGGGTCTGAGAGTCTGTGAAGATGGCCAGTCTTCTATCCCCACCTAAAAAGACC AAGCATCTCGAG		
	ORF Start: at 7		ORF Stop: at 703
	SEQ ID NO: 426	232 aa	MW at 27170.6kD
NOV55h, CG59595-05 Protein Sequence	DKRLRDNHEWKKLIMVQHWPE TVCEKIQDDCRDPPDYWTIHGLWPKSEGCNRSWPFN LEEIKDLLPEMRAYWPDVIHSFPNRSRFWKHEWEKHGTCAAQVDALNSQKKYFGRSLE LYRELDLNSVLLKLGIKPSINYQVADF KDALARVYGVI PKIQCLPPSRDEEVQTIGQ IELCLTKDQQLQNCTEPGEQSPKQEVWLANGAAESRGLRVCE DGPVFYPPPKTKH		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 55B.

Table 55B. Comparison of NOV55a against NOV55b through NOV55h.		
Protein Sequence	NOV55a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV55b	23..256 1..234	233/234 (99%) 234/234 (99%)
NOV55c	23..256 1..234	229/234 (97%) 231/234 (97%)
NOV55d	23..256 1..234	231/234 (98%) 234/234 (99%)
NOV55e	25..256 1..232	232/232 (100%) 232/232 (100%)
NOV55f	1..256 1..256	256/256 (100%) 256/256 (100%)
NOV55g	25..256 1..232	228/232 (98%) 229/232 (98%)
NOV55h	25..256 1..232	230/232 (99%) 232/232 (99%)

- 5 Further analysis of the NOV55a protein yielded the following properties shown in Table 55C.

Table 55C. Protein Sequence Properties NOV55a	
PSort analysis:	0.8200 probability located in outside; 0.1900 probability located in lysosome (lumen); 0.1000 probability located in endoplasmic reticulum (membrane); 0.1000 probability located in endoplasmic reticulum (lumen)

SignalP analysis:	Cleavage site between residues 25 and 26
-------------------	--

A search of the NOV55a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 55D.

Table 55D. Geneseq Results for NOV55a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV55a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAY21852	Human signal peptide-containing protein (SIGP) (clone ID 2652271) - Homo sapiens, 256 aa. [WO9933981-A2, 08-JUL-1999]	1..256 1..256	256/256 (100%) 256/256 (100%)	e-158
AAW75103	Human secreted protein encoded by gene 47. clone HMCBP63 - Homo sapiens, 256 aa. [WO9839446-A2, 11-SEP-1998]	1..256 1..256	256/256 (100%) 256/256 (100%)	e-158
AAY48563	Human breast tumour-associated protein 24 - Homo sapiens, 284 aa. [DE19813839-A1, 23-SEP-1999]	1..256 29..284	255/256 (99%) 255/256 (99%)	e-157
ABG12714	Novel human diagnostic protein #12705 - Homo sapiens, 342 aa. [WO200175067-A2, 11-OCT-2001]	1..256 85..342	247/258 (95%) 251/258 (96%)	e-150
ABG12711	Novel human diagnostic protein #12702 - Homo sapiens, 193 aa. [WO200175067-A2, 11-OCT-2001]	49..256 1..193	184/208 (88%) 187/208 (89%)	e-109

- 5 In a BLAST search of public sequence databases, the NOV55a protein was found to have homology to the proteins shown in the BLASTP data in Table 55E.

Table 55E. Public BLASTP Results for NOV55a				
Protein Accession Number	Protein/Organism/Length	NOV55a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
O00584	Ribonuclease 6 precursor (EC 3.1.27.-) - Homo sapiens (Human), 256 aa.	1..256 1..256	256/256 (100%) 256/256 (100%)	e-158

S78046	ribonuclease 6 (EC 3.1.27.-) precursor – human, 189 aa.	1..181 1..181	180/181 (99%) 180/181 (99%)	e-109
Q9CQ01	Ribonuclease 6 precursor (EC 3.1.27.-) - Mus musculus (Mouse), 259 aa.	1..256 1..259	176/261 (67%) 207/261 (78%)	e-105
JE0172	ribonuclease T2 (EC 3.1.27.1) - pig, 200 aa.	32..253 1..200	149/223 (66%) 172/223 (76%)	5e-88
JE0173	ribonuclease T2 (EC 3.1.27.1) - bovine, 198 aa.	33..250 2..196	126/219 (57%) 155/219 (70%)	2e-72

PFam analysis predicts that the NOV55a protein contains the domains shown in the Table 55F.

Table 55F. Domain Analysis of NOV55a			
Pfam Domain	NOV55a Match Region	Identities/ Similarities for the Matched Region	Expect Value
ribonuclease_T2	39..219	63/212 (30%) 149/212 (70%)	9.1e-64

### Example 56.

5 The NOV56 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 56A.

Table 56A. NOV56 Sequence Analysis			
	SEQ ID NO: 427	2684 bp	
NOV56a, CG92142-01 DNA Sequence	ATCAGAATTTTGAGTCTAGTATTTACTCTCTCGATTCCCTTGTTAATTTAAATGGGTAC CTATTTTATATAGCACATGATTGGGAATTACACTTTGTGACATGGATGAATCTGCAC TGACCCTTGGTACAATAGATGTTTCTTATCTGCCACATTCATCAGAATACAGTGTGG TCGATGTAAGCACACAAGTGAGGAATGGGTAGAGTGTGGCTTTAGACCCACCATCTTC AGATCTGCAACTTTAAAATGGAAAAGGAAAGCCTAATGAGTCGGAAGGCCATTTGTTG GAAGATGTTGTTACTCCTGCACTCCCCAGAGCTGGGACAAATTTTCAACCCAGTAT CCCGTCTTTGGGTTTGGGAATGTTATTTATATCAATGAAACTCACACAAGACACCCGC GGATGGCTTGCAAGACGCCTTTCTTACGTTCTTTTATTCAAGAGCGAGATGTGCATA AGGGCATGTTTGCCACCAATGTGACTGAAAATGTGCTGAACAGCAGTAGAGTACAAGA GGCAATTCAGAAAGTGGCTGCTGAATTAACCCCTGATGGTTCTGCCCAGCAGCAATCA AAAGCCGTTAACAAGTGAAAAAGAAAGCTAAAAGGATTCTTCAAGAAATGGTTGCCA CTGTCTCACCAGCAATGATCAGACTGACTGGGTGGGTGCTGCTAAAATGTTCAACAG CTTCTTTTGGAACTTCAAATTCACAAAGGTCAACTTGAGATGGTTAAAGCTGCAACT GAGACGAATTTGCCGCTTCTGTTTCTACCAAGTTCATAGATCCCATATTGACTATCTGC TGCTCACTTTCATTCTTCTGCTCCATAACATCAAAGCACCATACATTGCTTCAGGCAA TAATCTCAACATCCCAATCTTCAGTACCTTGATCCATAAGCTTGGGGGCTTCTTCATA CGACGAAGGCTCGATGAAACACCAGATGGACGGAAGATGTTCTCTATAGAGCTTTGC TCCATGGGCATATAGTTGAATTACTTCGACAGCAGCAATCTTGGAGATCTTCCTGGA AGGCACACGTTCTAGGAGTGGAAAAACCTCTGTGCTCGGGCAGGACTTTTGTGAGTT GTGGTAGATACTGTCTACCAATGTCTATCCAGACATCTTGATAATACCTGTTGGAA TCTCCTATGATCGCATTATCGAAGGTCACTACAATGGTGAACAACCTGGGCAACCTAA GAAGAATGAGAGCCTGTGGAGTGTAGCAAGAGGTGTTATTAGAATGTTACGAAAAAAC		

	TATGGTTGTGTCCGAGTGGATTTTGCACAGCCATTTTCCCTTAAAGGAATATTTAGAAA GCCAAAGTCAGAAACCGGTGTCTGCTCTACTTTCCCTGGAGCAAGCGTTGTTACCAGC TATACTTCCTTCAAGACCCAGTGATGCTGCTGATGAAGGTAGAGACACGTCCATTAAT GAGTCCAGAAATGCAACAGATGAATCCCTACGAAGGAGGTTGATTGCAAATCTGGCTG AGCATATTCTATTCAGTGTAGCAAGTCTGTGCCATTATGTCCACACACATTGTGGC TTGCCCTGCTCCTCTACAGACACAGGCAGGAATTGATCTCTCCACATTGGTCTGAAGAC TTCTTTGTGATGAAAGAGGAAGTCTGGCTCGTGATTTTGACCTGGGGTTCTCAGGAA ATTGAGAAGATGTAGTAATGCATGCCATACAGCTGCTGGGAAATTGTGTCACAATCAC CCACACTAGCAGGAACGATGAGTTTTTTATACCCCCAGCACAACTGTCCCATCAGTC TTCGAACTCAACTTCTACAGCAATGGGGTACTTCATGTCTTTATCATGGAGGCCATCA TAGCTTGCAGCCTTTATGCAGTTCTGAACAAGAGGGGACTGGGGGGTCCCCTAGCAC CCCACCTAACCTGATCAGCCAGGAGCAGCTGGTGCGGAAGCGGCCAGCTGTGCTAC CTTCTCTCCAATGAAGGCACCATCTCACTGCCTTGCCAGACATTTTACCAAGTCTGCC ATGAAACAGTAGGAAAGTTTATCCAGTATGGCATTCTTACAGTGGCAGAGCACGATGA CCAGGAAGATATCAGTCTAGTCTTGCTGAGCAGCAGTGGGACAAGAGCTTCTTGAA CCTTTGTCTTGGAGAAGTGATGAAGAAGATGAAGACAGTGACTTTGGGGAGGAACAGC GAGATTGCTACCTGAAGGTGAGCCAATCCAAGGAGCACCAGCAGTTTATCACCTTCTT ACAGAGACTCCTTGGGCCTTTGCTGGAGGCCTACAGCTCTGCTGCCATCTTTGTTCAC AACTTCAGTGGTCTGTTCAGAACCTGAGTATCTGCAAAAGTTGCACAAATACCTAA TAACCAGAACAGAAAGAAATGTTGCAGTATATGCTGAGAGTGCCACATATTGTCTTGT GAAGAATGCTGTGAAAATGTTTAAAGGATATTGGGGTTTTCAAGGAGACCAAAACAAAAG AGAGTGTCTGTTTTAGAACTGAGCAGCACTTTCTACCTCAATGCAACCGACAAAAAC TTCTAGAATATATTCTGAGTTTGTGGTGTCTGAGGTAACTGTGGCACTGCTGGCAA ATGAAGGTCATGAGATGAGTTCCTTGTAGGTACCAGCTTCTGGCTCAAGAGTTGAAGG TGCCGTCGCAGGGTCA		
	ORF Start: ATG at 101		ORF Stop: TAG at 2585
	SEQ ID NO: 428	828 aa	MW at 93835.7kD
NOV56a, CG92142-01 Protein Sequence	MDESALTGLGTIDVSYLPHSSEYSVGRCKHTSEEWVECGFRPTIFRSATLKWKESLMSR KRPFVGRCCYSCTPQSWDKFFNPISPSLGLRNVIIYNETHTRHRGLWARRLSYVLFIQ ERDVHKGMFATNVTENVLNSRRVQEAIAEVAELNPDGSAQQQSKAVNKVKKKAKRIIL QEMVATVSPAMIRLTGWVLLKLFNSFFWNIQIHKQLEMVKAATETNLPFLFVPHRS HIDYLLLTFFILFCHNIKAPYIASGNNLNPIPFSTLIHKLGGFFIRRRLDETPDGRKDV LYRALLHGHIVELLRQQQFLEIFLEGTRSRSGKTSCARAGLLSVVVDLTSTNVI PDIL IIPVGISYDRIIEGHYNGEQLGKPKKNESLWSVARGVIRMLRKNYGCVRVDFAPFSL KEYLESQSQKPVSAALLSLEQALLPAILPSRPSDADEGRDTSINESRNATDESLLRRRL IANLAEHILFTASKSCAIMSTHIVACLLLYRHRQIDLSTLVEDFFVMKEEVLARDFD LGFGSGNSEDVVMHAIQLLGNCVTITHTSRNDFFITPSTTVPSVFELNFYSNGVLHVF IMEAIIACSLYAVLNKRLGLGGPTSTPPNLISQEQLVKRAASLCYLLSNEGTISLPCQT FYQVCHETVGKFIQYGILTVAEHDDQEDISPLAEQQWDKLLPEPLSWRSDEEDEDSD FGEEQRDCYLKVSQSKEHQFFITFLQRLGLPLEAYSSAAIFVHNFGSPVPEPEYLQK LHKYLITRTERNVAVYAESATYCLVKNVAVKMKFDIGVFKETKQKRVSVLELSSTFLPQ CNRQKLLEYILSFVVL		
	SEQ ID NO: 429	2527 bp	
NOV56b, CG92142-02 DNA Sequence	GCACATGATTTGGGAATTACACTTTGTGACATGGATGAATCTGCACTGACCCTTGGTA CAATAGATGTTTCTTATCTGCCACATTATCAGAATACAGTGTGGTGCATGTAAGCA CACAAAGTGAGGAATGGGGTGAGTGTGGCTTTAGACCCACCGTCTTCAGATCTGCAACT TTAAATGGAAGAAAGCCATAAGTGCAGGAAAGGCCATTTGTTGGAAGATGTTGTT ACTCTGCATCCCAAGAGCTGGGACAAATTTTCAACCCAGTATCCCGTCTTTGGG TTTGCAGGAATGTTATTTATATCAATGAACTCACACAAGACACCGCGGATGGCTTGCA AGACGCCTTTCTACGTCTTTTTATTCAAGAGCGAGATGTGCATAAGGGCATGTTTG CCACCAATGTGACTGGAAATGTGCTGAACAGCAGTAGAGTACAAGAGGCAATTGCAGA AGTGGCTGCTGAATTAAACCTGATGGTTCTGCCAGCAGCAATCAAAGCCGTTAAC AAAGTGAAAAAGAAAGCTAAAAGGATTCTCAAGAAATGGTTGCCACTGTCTCACC GG CAATGATCAGACTGACTGGGTGGGTGCTGCTAAAACCTGTTCAACAGCTTCTTTTGGAA CATTCAAATTCACAAAGGTCAACTTGAGATGGTTAAAGCTGCAACTGAGACGAATTTG CCGCTTCTGTTTCTACCAGTTCATAGATCCCATATTGACTATCTGCTGCTCACTTTCA TTCTCTTCTGCCATAACATCAAAGCACCATAATTGCTTCAGGCAATAATCTCAACAT CCCAATCTTCAGTACCTTGATCCATAAGCTTGGGGGCTTCTTCATACGACGAAGGCTC		

	GATGAAACACCAGATGGACGGAAGATGTTCTCTATAGAGCTTTGCTCCATGGGCATA TAGTTGAATTACTTCGACAGCAGCAATTCTTGGAGATCTTCCCTGGAAGGCACACGTTT TAGGAGTGGAAAAACCTCTGTGCTCGGGCAGGACTTTTGTGCTGTTGGTAGATACT CTGTCTACCAATGTCATCCAGACATCTTGATAATACCTGTTGGAATCTCCTATGATC GCATTATCGAAGGTCACATAATGGTGAACAACCTGGGCAAACCTAAGAAGATGAGAG CCTGTGGAGTGTAGCAAGAGGTGTTATTAGAATGTTACGAAAAAATATGGTTGTGTC CGAGTGGATTTGACAGCCATTTTCTTAAAGGAATATTTAGAAAGCCAAAGTCAGA AACCGGTGCTGCTCTACTTTCCCTGGAGCAAGCGTTGTTACCAGCTATACTTCTTTC AAGACCCAGTGATGCTGCTGATGAAGGTAGAGACACGTCCATTAAATGAGTCCAGAAAT GCAACAGATGAATCCCTACGAAGGAGTTGATTGCAAATCTGGCTGAGCATATTCTAT TCACTGCTAGCAAGTCTGTGCCATTATGTCCACACATTGTGGCTTGCCTGCTCCT CTACAGACACAGGCAGGGAATTGATCTCTCCACATTGGTTCGAAGACTTCTTTGTGATG AAAGAGGAAGTCTGGCTCGTGATTTTGACCTGGGGTTCTCAGGAAATTCAGAAGATG TAGTAATGCATGCCATACAGCTGCTGGGAAATTGTGTCAATACCCACACTAGCAG GAATGATGAGTTTTTATCACCCCCAGCACAACTGTCCCATCAGTCTTCGAACTCAAC TTCTACAGCAATGGGGTACTTCATGTCTTTATCATGGAGGCCATCATAGCTTGCAGCC TTTATGCACTTCTGAACAAGAGGGGACTGGGGGGTCCCCTAGCACCCCACTAACCT GATCAGCCAGGAGCAGCTGGTGCGGAAGGCGGCCAGCCTGTGTACCTTCTCTCCAAT GAAGGCACCATCTCACTGCCTTGCCAGACATTTTACCAAGTCTGCCATGAAACAGTAG GAAAGTTTATCCAGTATGGCATTCTTACAGTGGCAGAGCAGATGACCAGGAAGATAT CAGTCTAGTCTTGCTGAGCAGCAGTGGGACAAGAAGCTTCTGAACCTTTGTCTTGG AGAAGTGATGAAGAAGATGAAGACAGTGACTTTGGGGAGGAACAGCGAGATTGCTACC TGAAGGTGAGCCAATCCAAGGAGCACCAGCAGTTTATCACCTTCTTACAGAGACTCCT TGGGCCTTTGCTGGAGGCTACAGCTCTGCTGCCATCTTTGTTCAACATTCAGTGGT CCTGTTCCAGAACCTGAGTATCTGCAAAAGTTGCACAAATACCTAATAACCAGAACAG AAAGAAATGTTGCAGTATATGCTGAGAGTGCCACATATTGTCTTGTGAAGAATGCTGT GAAAATGTTTAAGGATATTGGGGTTTCAAGGAGACCAACAAAAGAGAGTGTCTGTT TTAGAAGTGAAGCAGCACTTTTCTACCTCAATGCAACCGACAAGACTTCTAGAATATA TTCTGAGTTTGTGGTCTGTAAGTAACGTGTG		
	ORF Start: ATG at 31	ORF Stop: TAA at 2515	
	SEQ ID NO: 430	828 aa	MW at 93735.6kD
NOV56b, CG92142-02 Protein Sequence	MDESALTGLTIDVSYLPHSSEYSVGRCKHTSEEWGECGFRPTVFRSATLKWKESLMSR KRPFGVGRCCYSTPQSWDKFFNPSPSLGLRNVYINETHRHRGWLARRLSYVLFQ ERDVHKGMFATNVTGNVLNSSRVQEAIAEVAELNPDGSAQQSKAVNKVKKAKRIL QEMVATVSPAMIRLTGWVLLKLFNSFFWNIQIHKQLEMVKAATETNPLFLPVHRS HIDYLLTFILFCHNIKAPYIASGNLNIPIFSTLIHKLGGFFIRRLDETPDGRKDV LYRALLHGHIVELLRQQQFLEIFLEGTRSRSGKTSARAGLLSVVVDLTSTNVIPIIL IIPVGISYDRIIEGHYNGEQLGPKKNESLWSVARGVIRMLRKNYGCVRVDFAPFSL KEYLESQSQKPVSAALLSLEQALLPAILPSRPSDADEGRDTSINESRNATDESLRRRL IANLAEHILFTASKSCAIMSTHIVACLLLYRHRQIDLSTLVEDFFVMKEEVLARDFD LGFSGNSEDEVVMHAIQLLGNCVTITHTSRNDEFFITPSTTVPSVFELNFSYNGVLHVF IMEAIIACSLYAVLNKRGLGGPTSTPPNLISQQLVRKAASLCYLLSNEGTTISLPCQT FYQVCHETVGKFIQYGILTVAEHDDQEDISPSLAEQQWKKLPEPLSWRSDEEDEDSD FGEEQRDCYLKVSQSKEHQFFITFLQRLGLPLEAYSSAAIFVHNFSGPVPEPEYLQK LHKYLITRTERNVAVYAESATYCLVKNVAMFKDIGVFKETKQKRVSVLELSSTFLPQ CNRQRLLEYILSFVVL		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 56B.

Table 56B. Comparison of NOV56a against NOV56b.		
Protein Sequence	NOV56a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV56b	1..828	824/828 (99%)
	1..828	826/828 (99%)

Further analysis of the NOV56a protein yielded the following properties shown in Table 56C.

<b>Table 56C. Protein Sequence Properties NOV56a</b>	
PSort analysis:	0.8500 probability located in endoplasmic reticulum (membrane); 0.4400 probability located in plasma membrane; 0.3000 probability located in nucleus; 0.1000 probability located in mitochondrial inner membrane
SignalP analysis:	No Known Signal Sequence Predicted

5 A search of the NOV56a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 56D.

<b>Table 56D. Geneseq Results for NOV56a</b>				
<b>Geneseq Identifier</b>	<b>Protein/Organism/Length [Patent #, Date]</b>	<b>NOV56a Residues/ Match Residues</b>	<b>Identities/ Similarities for the Matched Region</b>	<b>Expect Value</b>
ABG66665	Human glycerol-3-phosphate acyltransferase hGPAT - Homo sapiens, 828 aa. [WO200240666-A2, 23-MAY-2002]	1..828 1..828	827/828 (99%) 827/828 (99%)	0.0
AAE22144	Human TRNFR-6 protein - Homo sapiens, 828 aa. [WO200226950-A2, 04-APR-2002]	1..828 1..828	827/828 (99%) 827/828 (99%)	0.0
AAU78393	Human acyltransferase, ACTR-1 - Homo sapiens, 828 aa. [WO200216592-A2, 28-FEB-2002]	1..828 1..828	826/828 (99%) 827/828 (99%)	0.0
AAE22145	Human TRNFR-7 protein - Homo sapiens, 801 aa. [WO200226950-A2, 04-APR-2002]	56..826 40..799	262/790 (33%) 403/790 (50%)	e-102
ABB61594	Drosophila melanogaster polypeptide SEQ ID NO. 11574 - Drosophila melanogaster, 850 aa. [WO200171042-A2, 27-SEP-2001]	163..809 194..820	196/654 (29%) 353/654 (53%)	4e-82

In a BLAST search of public sequence databases, the NOV56a protein was found to have homology to the proteins shown in the BLASTP data in Table 56E.

<b>Table 56E. Public BLASTP Results for NOV56a</b>				
<b>Protein Accession</b>	<b>Protein/Organism/Length</b>	<b>NOV56a Residues/</b>	<b>Identities/ Similarities for</b>	<b>Expect Value</b>

Number		Match Residues	the Matched Portion	
Q9HCL2	Glycerol-3-phosphate acyltransferase, mitochondrial precursor (EC 2.3.1.15) (GPAT) - Homo sapiens (Human), 828 aa.	1..828 1..828	828/828 (100%) 828/828 (100%)	0.0
AAH30783	KIAA1560 protein - Homo sapiens (Human), 828 aa.	1..828 1..828	825/828 (99%) 825/828 (99%)	0.0
Q8VCT2	Glycerol-3-phosphate acyltransferase, mitochondrial - Mus musculus (Mouse), 827 aa.	1..828 1..827	769/828 (92%) 799/828 (95%)	0.0
Q61586	Glycerol-3-phosphate acyltransferase, mitochondrial precursor (EC 2.3.1.15) (GPAT) (P90) - Mus musculus (Mouse), 827 aa.	1..828 1..827	767/828 (92%) 799/828 (95%)	0.0
P97564	Glycerol-3-phosphate acyltransferase, mitochondrial precursor (EC 2.3.1.15) (GPAT) - Rattus norvegicus (Rat), 828 aa.	1..828 1..828	760/828 (91%) 794/828 (95%)	0.0

PFam analysis predicts that the NOV56a protein contains the domains shown in the Table 56F.

Table 56F. Domain Analysis of NOV56a			
Pfam Domain	NOV56a Match Region	Identities/ Similarities for the Matched Region	Expect Value
Acyltransferase	215..412	47/207 (23%) 151/207 (73%)	6.4e-34

### Example 57.

5. The NOV57 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 57A.

Table 57A. NOV57 Sequence Analysis			
	SEQ ID NO: 431	1538 bp	
NOV57a, CG95765-01. DNA Sequence	CACCGAGCCTCACGGGAGCTGATGGCTGCAAAGAAGACCCACACCTCACAAATTGAAG TGATCCCTTGCAAATCTGTGGGGACAAGTCGTCTGGGATCCACTACGGGGTTATCAC CTGTGAGGGGTGCAAGGGCTTCTCCGGCCTACTCCTGCACCCGTCAGCAGAACTGCC CCATCGACCGCACCAGCCGAAACCGATGCCAGCACTGCCGCCTGCAGAAATGCCTGGC GCTGGGGATGTCCGAGATGCTGTCAAGTTCGGCCGCATGTCCAAGAAGCAGAGGGAC AGCCTGCATGCAGAAAGTGCAGAAACAGCTGCAGCAGCGGCAACAGCAGCAACAGGAAC CAGTGGTCAAGACCCCTCCAGCAGGGGCCCAAGGAGCAGATACCCTCACCTACACCTT		

	GGGGCTCCCAGACGGGCAGCTGCCCCCTGGGCTCCTCGCCTGACCTGCCTGAGGCTTCT GCCTGTCCCCCTGGCCTCCTGAAAGCCTCAGGCTCTGGGCCCTCATATTCCAACAACT TGGCCAAGGCAGGGCTCAATGGGGCCTCATGCCACCTTGAATACAGCCCTGAGCGGGG CAAGGCTGAGGGGAGAGAGAGCTTCTATAGCACAGGCAGCCAGCTGACCCCTGACCGA TGTGGACTTCGTTTTGAGGAACACAGGCATCCTGGGCTTGGGGAAGTGGGACAGGGCC CAGACAGCTACGGCAGCCCCAGTTTCCGCAGCACACCGGAGGCACCTATGCCCTCCCT GACAGAGATAGAGCACCTGGTGCAGAGCGTCTGCAAGTCTACAGGGAGACATGCCAG CTGCGGCTGGAGGACCTGCTGCGGCAGCGCTCCAACATCTTCTCCCGGGAGGAAGTGA CTGGCTACCAGAGGAAGTCCATGTGGGAGATGTGGGAACGGTGTGCCACCACCTCAC CGAGGCCATTAGTACGTGGTGGAGTTTCGCCAAGAGGCTCTCAGGCTTTATGGAGCTC TGCCAGAATGACCAGATTGTGCTTCTCAAAGCAGGAGCAATGGAAGTGGTGTGGTTA GGATGTGCCGGGCTACAATGCCAACACACACAGTCTTTTTTGAAGGCAAAATACGG TGGTGTGGAGCTGTTTTCGAGCCTTGGGCTGCAGCGAGCTCATCAGCTCCATATTTGAC TTTTCCCACTTCTCTCAGCGCCTGTGTTTTTCCGAGGATGAGATTGCCCTCTACACGG CCCTTGTTCTCATCAATGCCAACCGTCTCTGGGCTCCAAGAGAAGAGGAGAGTGGAA TCTGCAATACAATTTGGAAGTGGCTTTCCATCATCATCTCTGCAAGACTCATCGACAA AGCATCTGGCAAAGCTGCCACCCAAAGGAAAACCTCCGAGCGTGTGTAGCCAGCATG TGGAAAGGCTGCAGATCTTCCAGCACCTCCACCCCATCGTGGTCCAAGCCGCTTTCCC TCCACTCTACAAGGAGCTTTCAGCACTGAAACCGAGTCACCTGTGGGGCTGTCCAAG TGACCTGGAAGAGGGAATCCTTGCCCTCTCC		
	ORF Start: ATG at 240		ORF Stop: TGA at 1509
	SEQ ID NO: 432	423 aa	MW at 47418.4kD
NOV57a, CG95765-01 Protein Sequence	MSRDAVKFGRMSKKQRDSLHAEVQKQLQORQQQQEPVVKTPPAGAQAADTLTYTLGL PDGQLPLGSSPDLPEASACPPGLLKASGSGPSYSNNLAKAGLNGASCHLEYSPERGKA EGRESFYSTGSQLTPDRCLRFEEHRHPGLGELGQGPDSYSGSPFRSTPEAPYASLTE IEHLVQSVCKSYRETCQLRLEDLLRQRSNIFSRREVTGYQRKSMWEMWERCAHHLTEA IQYVVEFAKRLSGFMELCQNDQIVLLKAGAMEVVLVRMCRAYNANNHTVFEGKYGGV ELFRALGCSELISSIFDFSHFLSALCFSEDEIALYTALVLINANRPGLEKRRVEHLQ YNLELAFHHHLCKTHRQSLAKLPPKGLRSLCSQHVERLQIFQHLHPVVQAAFPPL YKELFSTETESPVGLSK		
	SEQ ID NO: 433	1819 bp	
NOV57b, CG95765-02 DNA Sequence	CCCCTGGGCCCTGCTCCCTGCCCTCCTGGGCAGCCAGGGCAGCCAGGACGGCACCAAG GGAGCTGCCCCATGGACAGGGCCCCACAGAGACAGCACCGAGCCTCACGGGAGCTGCT GGCTGCAAAGAAGACCCACACCTCACAATTTGAAGTGATCCCTTGCAAAATCTGTGGG GACAAGTCGTCTGGGATCCACTACGGGGTTATCACTGTGAGGGGTGCAAGGGCTTCT TCCGCCGAGCCAGCGCTGTAACGCGGCCTACTCCTGCACCCGTGAGCAGATGCTCC CATCGACCGCACCAGCCGAAACCGATGCCAGCACTGCCGCTGCAGAAATGCCTGGCG CTGGGGATGTCCCGAGATGCTGTCAAGTTTCGGCCGCATGTCCAAGAAGCAGAGGGACA GCCTGCATGCAGAAGTGCAGAAACAGCTGCAGCAGCGGCAACAGCAGCAACAGGAACC AGTGGTCAAGACCCCTCCAGCAGGGGCCAAGGAGCAGATACCTCACCTACACCTTG GGGCTCCCAGACGGGCAGCTGCCCTGGGCTCCTCGCCTGACCTGCCTGAGGCTTCTG CCTGTCCCCCTGGCCTCCTGAAAGCCTCAGGCTCTGGGCCCTCATATTCCAACAACCTT GGCCAAGGCAGGGCTCAATGGGGCCTCATGCCACCTTGAATACAGCCCTGAGCGGGGC AAGGCTGAGGGCAGAGAGAGCTTCTATAGCACAGGCAGCCAGCTGACCCCTGACCGAT GTGGACTTCGTTTTGAGGAACACAGGCATCCTGGGCTTGGGGAAGTGGGACAGGGCCC AGACAGCTACGGCAGCCCCAGTTTCCGCAGCACACCGGAGGCACCTATGCCTCCCTG ACAGAGATAGAGCACCTGGTGCAGAGCGTCTGCAAGTCTACAGGGAGACATGCCAGC TGCGGCTGGAGGACCTGCTGCGGCAGCGCTCCAACATCTTCTCCCGGAGGAAGTGAC TGGCTACCAGAGGAAGTCCATGTGGGAGATGTGGGAACGGTGTGCCACCACCTCACC GAGGCCATTAGTACGTGGTGGAGTTCGCCAAGAGGCTCTCAGGCTTTATGGAGCTCT GCCAGAATGACCAATTGTGCTTCTCAAAGCAGGAGCAATGGAAGTGGTGTGTTAGT GATGTGCCGGGCTACAATGCTGACAACCGCACGGTCTTTTTTGAAGGCAAAATACGGT GGCATGGAGCTGTTCCGAGCCTTGGGCTGCAGCGAGCTCATCAGCTCCATCTTTGACT TCTCCCACTCCCTAAGTGCCTTGCACTTTTCCGAGGATGAGATTGCCCTCTACACAGC CCTTGTTCTCATCAATGCCCATCGGCCAGGGCTCCAAGAGAAAAGGAAAGTAGAACAG CTGCAGTACAATCTGGAGCTGGCCTTTCATCATCATCTCTGCAAGACTCATCGCCAAA GCATCCTGGCAAAGCTGCCACCCAAGGGGAAGCTTCGAGCCTGTGTAGCCAGCATGT GGAAAGGCTGCAGATCTTCCAGCACCTCCACCCCATCGTGGTCCAAGCCGCTTTCCCT		

	CCACTCTACAAGGAGCTCTTCAGCACTGAAACCGAGTCACCTGTGGGCTGTCCAAGTG ACCTGGAAGAGGGACTCCTTGCCCTCTCCCTATGGCCTGTGGCCACCTCCCTGGACCC CGTTCCACCCTCACCTTTTCCCTTCCCATGAACCTGGAGGGTGGTCCCCACCAGCT CTTTGGAAGTGAGCAGATGCTGCGCTGGCTTTCTGTGACAGGCCGGCCTGGCAGTG GGACAATCGCCAGAGGGTGGG		
	ORF Start: ATG at 70		ORF Stop: TGA at 1750
	SEQ ID NO: 434	560 aa	MW at 62588.6kD
NOV57b, CG95765-02 Protein Sequence	MDRAPQRQHRASRELLAAKKTHTSQIEVIPCKICGDKSSGIHYGVITCEGCKGFFRRS QRCNAAYSCTRQQNCPIDRTSRNCQHCLQKCLALGMSRDAVKFGRMSKKQRDSLHA EVQKQLQQRQQQQEPVVKTPPAGAQQADTLTYTLGLPDGQLPLGSSPDLPASACPP GLLKASGSGPSYSNNLAKAGLNGASCHLEYS PERGKAEGRESFYSTGSQLTDPDRCGLR FEEHRHPGLGELGQGPSYGSFSRSTPEAPYASLTEIHLVQSVCKSYRETCQLRLE DLLRQRSNIFSRREVTGYQRKSMWEMWERCAHHLTEAIQYVVEFAKRLSGFMELCQND QIVLLKAGAMEVVLVRMCRAYNADNRTVFFEGKYGGMELFRALGCSELISSIFDFSHS LSALHFSEDEIALYALVLI NAHRPGLQEKRKVEQLQYNLELAFHHHLCKTHRQSILA KLPPKGKLRSLCSQHVERLQIFQHLHPIVVQAAFPPLYKELFSTETESPVGCPDLEE GLLASPYGLLATSLDPVPPSPFSFPMNPGGWSPPALWK		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 57B.

Table 57B. Comparison of NOV57a against NOV57b.		
Protein Sequence	NOV57a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV57b	1..420 96..515	412/420 (98%) 416/420 (98%)

- 5 Further analysis of the NOV57a protein yielded the following properties shown in Table 57C.

Table 57C. Protein Sequence Properties NOV57a	
PSort analysis:	0.3600 probability located in mitochondrial matrix space; 0.3000 probability located in microbody (peroxisome); 0.1000 probability located in lysosome (lumen); 0.0000 probability located in endoplasmic reticulum (membrane)
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV57a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 57D.

Table 57D. Geneseq Results for NOV57a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV57a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value

AAB03062	Human retinoid-like orphan receptor-gamma 60 kD isoform - Homo sapiens, 518 aa. [WO200024757-A1, 04-MAY-2000]	1..423 96..518	415/423 (98%) 419/423 (98%)	0.0
AAB03066	Human ROR-gamma 60 kD isoform polymorphic variant #1, L516I - Homo sapiens, 518 aa. [WO200024757-A1, 04-MAY-2000]	1..423 96..518	414/423 (97%) 419/423 (98%)	0.0
AAB03069	Human ROR-gamma 60 kD isoform polymorphic variant #3, K518R - Homo sapiens, 518 aa. [WO200024757-A1, 04-MAY-2000]	1..423 96..518	414/423 (97%) 419/423 (98%)	0.0
AAB03068	Human ROR-gamma 60 kD isoform polymorphic variant #2 - Homo sapiens, 518 aa. [WO200024757-A1, 04-MAY-2000]	1..423 96..518	414/423 (97%) 419/423 (98%)	0.0
AAB03067	Human ROR-gamma 60 kD isoform polymorphic variant #1, L516V - Homo sapiens, 518 aa. [WO200024757-A1, 04-MAY-2000]	1..423 96..518	414/423 (97%) 419/423 (98%)	0.0

In a BLAST search of public sequence databases, the NOV57a protein was found to have homology to the proteins shown in the BLASTP data in Table 57E.

Table 57E. Public BLASTP Results for NOV57a				
Protein Accession Number	Protein/Organism/Length	NOV57a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
CAD38900	Hypothetical protein - Homo sapiens (Human), 497 aa.	1..423 75..497	415/423 (98%) 419/423 (98%)	0.0
AAH31554	Hypothetical protein - Homo sapiens (Human), 518 aa.	1..423 96..518	415/423 (98%) 419/423 (98%)	0.0
P51449	Nuclear receptor ROR-gamma (Nuclear receptor RZR-gamma) - Homo sapiens (Human), 560 aa.	1..420 96..515	412/420 (98%) 416/420 (98%)	0.0
Q91YT5	RAR-related orphan receptor gamma - Mus musculus (Mouse), 516 aa.	1..423 96..516	378/423 (89%) 395/423 (93%)	0.0

Q9R177	RORgamma t - Mus musculus (Mouse), 495. aa.	1..423 75..495	378/423 (89%) 395/423 (93%)	0.0
--------	--	-------------------	--------------------------------	-----

PFam analysis predicts that the NOV57a protein contains the domains shown in the Table 57F.

Table 57F. Domain Analysis of NOV57a			
Pfam Domain	NOV57a Match Region	Identities/ Similarities for the Matched Region	Expect Value
hormone_rec	230..411	56/210 (27%) 138/210 (66%)	1.1e-34

### Example 58.

The NOV58 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 58A.

Table 58A. NOV58 Sequence Analysis			
	SEQ. ID NO: 435	1712 bp	
NOV58a, CG97178-01. DNA Sequence	AAGGTCAATGATAGCATCTGCCTAGAGTCAAACCTCCGTGCTTCTCAGACAGTGCCTT TTCACCATGAGTGGGTGCCCATTTTTAGGAAACAACCTTGGATATACTTTTAAAAAAC TCCCCGTAGAAGGCAGCGAAGAAGACAAATCACAACTGGTGTGAATAGAGCCAGCAA AGGAGGTCTTATCTATGGGAACTACCTGCATTTGGAAAAAGTTTGAATGCACAAGAA CTGCAAAGTGAAACAAAAGGAAATAAAATCCATGATGAACATCTTTTATCATAAATC ATCAAGCTTATGAACTCTGGTTTAAGCAAATCCTCTGGGAGTTGGATTCTGTTTCGAGA GATCTTTCAGAAATGGCCATGTGAGAGATGAAAGGAACATGCTTAAGGTTGTTTCTCGG ATGCACCGAGTGTCTAGTGATCCTGAAACTGCTGGTGCAGCAGTTTCCATTCTGGAGA CGATGACAGCCTTGGACTTCAATGACTTCAGAGAGTACTTATCTCCAGCATCAGGCTT CCAGAGTTTGCAATTCCGACTATTAGAAAACAAGATAGGTGTTCTTCAGAACATGAGA GTCCCTTATAACAGAAGACATTATCGTGATAACTTCAAAGGAGAAGAAAATGAAGTGC TACTTAAATCTGAGCAGGAAAAGACACTTCTGGAATTAGTGGAGGCAATGGCTGGAAG AACTCCAGGTTTAGAGCCACATGGATTTAACTTCTGGGAAAGCTTGAAAAAATATC ACCAGAGGCCTGGAAGAGGAATTATCAAGGATTCAGGCTAAAGAAGAGTCTGAAGAAA AAGAGGAACAGGTGGCTGAATTTGAGAAGCAAAAAGAGGTGCTACTGCTTATTGGA TGAGAAACGTCATGAACATCTCCTTAGTAAAGGTGAAAGACGGCTGTCATACAGAGCA CTTCAGGGAGCATTGATGATATATTTTACAGGGAAGAGCCTAGGTTCCAGGTGCCTT TTCAGTTGCTGACTTCTCTTATGGACATAGATTCACTGATGACCAATGGAGATATAA CCATGTGTGCATGGTGCACAGAATGCTGGGCAGCAAAGCTGGCACCAGTGGTTCCTCA GGCTATCACTACCTGCGATCACTGTGAGTGATAGGTACAAGGTATTGTAGATTTAT TTAATCTTTCAACATACCTGATTCCCCGACACTGGATACCGAAGATGAACCAACCAT TCACAAATTTCTATATACAGCAGAATACTGTGATAGCTCCTACTTCAGCAGTGATGAA TCAGATTAAATCGTCTGCAAAATCTATGAAGAATACTGGTTTCACAGCCTATTTTTF ATTTTCTATGGATTTTCATAAATACAGTTTGAATATATGTATGCATATATTGTTTCAGC ACCACGATGCTCTGATTAAATCTAGAAAACAATTTGATTACCTCTTGTGTTGTGACAAG ACTAAGCATTAAGATGAGAAAGAATAACATTAAATAGTAACATTGTACATAGGGTGT TTCCTATTAAAAATCAGTTTCCCTGAGACTTAATGTAACTTAATGTAACTTAATGTAACT TCTCATGTTTCATCTTTATAAACTTGTAACCTTCATCTATTTCAAATATTTTATGCA GTACATTATATTATCTGTACAAAGGCTTCAAACAAAATTTTAAATAATAAAGTA TTAATCTTTCAAAAAAAAAAAAAAAAAAAAAA		
	ORF Start: ATG at 65		ORF Stop: TAA at 1283

	SEQ ID NO: 436	406 aa	MW at 47871.1kD
NOV58a, CG97178-01 Protein Sequence	MSGCPFLGNNFGYTFKKLPVEGSEEDKSQTVNRSKGGLIYGNYLHLEKVLNAQELQ SETKGNKIHDEHLFIITHQAYELWFKQILWELDSVREIFQNGHVRDERNMLKVVS RMSH RVSVILKLLVQQFSILETMTALDFNDFREYLSPASGFQSLQFRLLNKGIVLQNM RVP YNRRHYRDNFKGEENELLKSEQEKTLLELVEAWLERTPGLEPHGFNFWGKLEKNIT R GLEEEFIRIQAKEBSEEKEEQVAEFQKQKEVLLSLFDEKRHEHLLSKGERRLSY RALQ GALMIYFYREEPRFQVPFQLLTSMDIDSLMTKWRYNHVCMVHRLGSKAGTGGSS GY HYLRSTVSDRYKVFVDLFNLSTYLIPRHWIPKMNPITHKFLYTAEYCDSSYFSS DESD		
	SEQ ID NO: 437	1298 bp	
NOV58b, CG97178-02 DNA Sequence	GTGCTTCTCAGACAGTGCCTTTTACCACATGAGTGGGTGCCCATTTT TAGGAAACA ACTTTGGATATACTTTAAAAAACTCCCGTAGAAGGCAGCGAAGAAGACAAATC ACAAAC TGGTGTGAATAGAGCCAGCAAAGGAGGTCTTATCTATGGGAACCTGCTGATTT GGAA AAAGTTTTGAATGCACAAGAACTGCAAAGTGAAACAAAAGGAAATAAAATCCAT GATG AACATCTTTTTATCATACTCATCAAGCTTATGAACTCTGGTTTAAAGCAAATC CTCTG GGAGTTGGATTCTGTTTCGAGAGATCTTTCAGAATGGCCATGTCAGAGATGAA AGGAAC ATGCTTAAGGTGTGTTCTCGGATGCACCGAGTGTCACTGATCCTGAAACTGCT GGTGC AGCAGTTTTCCATTCTGGAGACGATGACAGCCTTGGACTTCAATGACTTCAGAG AGTA CTTATCTCCAGCATCAGGCTTCCAGAGTTTGCAATTCCGACTATTAGAAAACA AGATA GGTGTCTTTCAGAACATGAGAGTCCCTTATAACAGAAGACATTATCGTGATA ACTTCA AAGGAGAAGAAAATGAAGTCTACTTAAATCTGAGCAGGAAAAGACACTTCTG GAATT AGTGGAGGCATGGCTGGAAGAACTCCAGGTTTAGAGTCAATGGATTAACTTCT GCG GGAAAGCTTGAAAAAATATCACCAGAGGCCTGGAAGAGGAATTCATAAGGAT TCAGG CTAAAGAAGAGTCTGAAGAAAAAGAGGAACAGGTGGCTGAATTTCAAGCA AAAAAG GGTGCTACTGTCTTATTTGATGAGAAACGTCATGAACATCTCTTAGTAAAG GTGAA AGACGGCTGTCTACACAGAGCACTTCAGGGAGCATTGATGATATATTTTAC AGGGAAG AGCCTAGGTTCCAGGTGCCTTTTCAGTTGCTGACTTCTCTTATGGACATAG ATTCACT GATGACCAAATGGAGATATAACCATGTGTGCATGGTGCACAGAATGCTGGC AGCAAA GCTGGCACCGGTGGTTCCTCAGGCTATCACTACCTGCGATCAAACTGTGAG TATAGGT ACAAGGTATTTGTAGATTTATTTAATCTTTCAACATACCTGATTCCCGGAC ACTGGAT ACCGAAGATGAACCAACCATTCACAAATTTCTATATACAGCAGAACTACTG TGATAGC TCCTACTTCAGCAGTGATGAATCAGATTAATCGTCTGCAAAATCTATGAAG AATAC TGGTTTCACAGCCTATTTAAGG		
	ORF Start: ATG at 28		ORF Stop: TAA at 1246
	SEQ ID NO: 438	406 aa	MW at 47861.1kD
NOV58b, CG97178-02 Protein Sequence	MSGCPFLGNNFGYTFKKLPVEGSEEDKSQTVNRSKGGLIYGNYLHLEKVLNAQELQ SETKGNKIHDEHLFIITHQAYELWFKQILWELDSVREIFQNGHVRDERNMLKVVS RMSH RVSVILKLLVQQFSILETMTALDFNDFREYLSPASGFQSLQFRLLNKGIVLQNM RVP YNRRHYRDNFKGEENELLKSEQEKTLLELVEAWLERTPGLESHGFNFWGKLEKNIT R GLEEEFIRIQAKEBSEEKEEQVAEFQKQKEVLLSLFDEKRHEHLLSKGERRLSY RALQ GALMIYFYREEPRFQVPFQLLTSMDIDSLMTKWRYNHVCMVHRLGSKAGTGGSS GY HYLRSTVSDRYKVFVDLFNLSTYLIPRHWIPKMNPITHKFLYTAEYCDSSYFSS DESD		
	SEQ ID NO: 439	1240 bp	
NOV58c, 275481043 DNA Sequence	GCCGGATCCACCATGAGTGGGTGCCCATTTT TAGGAAACA ACTTTGGATATACTTTTAAAACTCCCGTAGAAGGCAGCGAAGAAGACAAATC ACAAACTGGTGTGAATAGAGC CAGCAAAGGAGGTCTTATCTATGGGAACCTGCATTGGAAAAAGTTTGAATGCA CAAGAACTGCAAAGTGAAACAAAAGGAAATAAAATCCATGATGAACATCTTTTATCA TAACTCATCAAGCTTATGAACTCTGGTTTAAAGCAAATCCTCTGGGAGTTGGATTCTGT TCGAGAGATCTTTCAGATGGCCATGTGAGAGATGAAAGGAACATGCTTAAGGTTGTT TCTCGGATGCACCGAGTGTCACTGATCCTGAAACTGCTGGTGCAGCAGTTTTCCATTC TGGAGACGATGACAGCCTTGGACTTCAATGACTTCAGAGAGTACTTATCTCCAGCATC AGGCTTCCAGAGTTTGAATCCGACTATTAGAAAACAAGATAGGTGTTCTTCAGAAC ATGAGAGTCCCTTATAACAGAGACATTAATCGTGATAACTTCAAAGGAGAAGAAAATG AACTGCTACTTAAATCTGAGCAGGAAAAGACACTTCTGGAATTAGTGGAGGCATGGCT GGAAAGAACTCCAGGTTTAGAGTCACATGGATTTAACTTCTGGGGAAAGCTTGAAAAA AATATCACCAGAGGCCTGGAAGAGGAATTCATAAGGATTCAGGCTAAAGAAGAGTCTG AAGAAAAAGAGGAACAGGTGGCTGAATTTCAGAAGCAAAAAGAGGTGCTACTGTCTT ATTTGATGAGAAACGTCATGAACATCTCTTAGTAAAGGTGAAAGACGGCTGTCTATAC		

	AGAGCACTTCAGGGAGCATTGATGATATATTTTACAGGGAAGAGCCTAGGTTCCAGG TGCCCTTTTCAGTTGCTGACTTCTCTTATGGACATAGATTCACTGATGACCAAATGGAG ATATAACCATGTGTGCATGGTGCACAGAATGCTGGGCAGCAAAGCTGGCACCAGGTGGT TCCTCAGGCTATCACTACCTGCGATCAACTGTGAGTGATAGGTACAAGGTATTTGTAG ATTTATTTAATCTTTCAACATACCTGATTCCCGACACTGGATACTGAAGATGAACCC AACCATTCAAAATTTCTATATACAGCAGAATACTGTGATAGCTCCTACTTCAGCAGT GATGAATCAGATGTCGACGGTG		
	ORF Start: at 1		ORF Stop: end of sequence
	SEQ ID NO: 440	414 aa	MW at 48464.7kD
NOV58c, 275481043 Protein Sequence	AGSTMSGCPFLGNNFGYTFKKLPVEGSEEDKSQTGVNRASKGGLIYGNYLHLEKVLNA QELQSETKGNKIHDEHLFIITHQAYELWFKQILWELDSVREIFQNGHVRDERNMLKV SRMHRVSVILKLLVQQFSILETMTALDFNDFREYLSPASGFQSLQFRLEENKIGVLQ MRVPYNRRHYRDNFKGEENELLLKSEQEKTLLELVEAWLERTPGLESHGFNFWGKLEK NITRGLLEEFIRIQAKEESEKKEEQVAEFQKQKEVLLSLFDEKRHEHLLSKGERRLSY RALQGALMIYFYREEPRFQVPFQLLTSLMDIDSLMTKWRYNHVCMVHRLGSKAGTGG SSGYHYLRSTVSDRYKVFVDLNLSTYLI PRHWILKMNPTIHKFLYTAEYCDSSYFSS DESDVDGX		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 58B.

Table 58B. Comparison of NOV58a against NOV58b and NOV58c.		
Protein Sequence	NOV58a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV58b	1..406	405/406 (99%)
	1..406	405/406 (99%)
NOV58c	1..406	404/406 (99%)
	5..410	404/406 (99%)

5 Further analysis of the NOV58a protein yielded the following properties shown in Table 58C.

Table 58C. Protein Sequence Properties NOV58a	
PSort analysis:	0.5095 probability located in microbody (peroxisome); 0.4500 probability located in cytoplasm; 0.1000 probability located in mitochondrial matrix space; 0.1000 probability located in lysosome (lumen)
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV58a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 58D.

Table 58D. Geneseq Results for NOV58a				
Geneseq	Protein/Organism/Length   Patent	NOV58a	Identities/	Exact

Identifier	#, Date]	Residues/ Match Residues	Similarities for the Matched Region	Value
AAR21549	Human Tryptophan Oxygenase TDO2 - Homo sapiens, 406 aa. [WO9202637-A, 20-FEB-1992]	1..406 1..406	403/406 (99%) 404/406 (99%)	0.0
AAR21547	Human Tryptophan-2,3-dioxygenase deduced from clone HTO3 - Homo sapiens, 436 aa. [WO9202637-A, 20-FEB-1992]	1..396 1..394	365/396 (92%) 369/396 (93%)	0.0
AAR21546	Human Tryptophan-2,3-dioxygenase deduced from clone HTO3 - Homo sapiens, 238 aa. [WO9202637-A, 20-FEB-1992]	1..228 1..228	225/228 (98%) 226/228 (98%)	e-130
ABB58903	Drosophila melanogaster polypeptide SEQ ID NO 3501 - Drosophila melanogaster, 379 aa. [WO200171042-A2, 27-SEP-2001]	19..389 4..374	213/373 (57%) 273/373 (73%)	e-115
AAU11269	Human N-acetyltransferase 1 (NAT1) variant polypeptide - Homo sapiens, 290 aa. [WO200179551- A1, 25-OCT-2001]	132..223 194..288	32/96 (33%) 44/96 (45%)	0.44

In a BLAST search of public sequence databases, the NOV58a protein was found to have homology to the proteins shown in the BLASTP data in Table 58E.

Table 58E. Public BLASTP Results for NOV58a				
Protein Accession Number	Protein/Organism/Length	NOV58a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
P48775	Tryptophan 2,3-dioxygenase (EC 1.13.11.11) (Tryptophan pyrrolase) (Tryptophanase) (Tryptophan oxygenase) (Tryptamin 2,3- dioxygenase) (TRPO) - Homo sapiens (Human), 406 aa.	1..406 1..406	406/406 (100%) 406/406 (100%)	0.0
Q8VCW3	Tryptophan 2,3-dioxygenase - Mus musculus (Mouse), 406 aa.	1..406 1..406	360/406 (88%) 388/406 (94%)	0.0
P48776	Tryptophan 2,3-dioxygenase (EC 1.13.11.11) (Tryptophan pyrrolase) (Tryptophanase) (Tryptophan oxygenase) (Tryptamin 2,3- dioxygenase) (TRPO) - Mus	1..406 1..406	359/406 (88%) 388/406 (95%)	0.0

	musculus (Mouse), 406 aa.			
P21643	Tryptophan 2,3-dioxygenase (EC 1.13.11.11) (Tryptophan pyrrolase) (Tryptophanase) (Tryptophan oxygenase) (Tryptamin 2,3-dioxygenase) (TRPO) - Rattus norvegicus (Rat), 406 aa.	1..406 1..406	360/406 (88%) 389/406 (95%)	0.0
O17440	VERMILION - Drosophila ananassae (Fruit fly), 380 aa.	19..389 4..375	214/374 (57%) 275/374 (73%)	e-115

PFam analysis predicts that the NOV58a protein contains the domains shown in the Table 58F.

Table 58F. Domain Analysis of NOV58a			
Pfam Domain	NOV58a Match Region	Identities/ Similarities for the Matched Region	Expect Value
No Significant Matches Found			

### Example 59.

The NOV59 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 59A.

5

Table 59A. NOV59 Sequence Analysis			
	SEQ ID NO: 441	1060 bp	
NOV59a, CG98102-01 DNA Sequence	CGCGGGCCGACTGGTGTTCATCCGTCACCTCGCCGAGGTTCCCTTGGGTCATGGTGCCAG CCTGACTGAGAAGAGACGCTCCCGGAGACGAATGAGGAACCACTCCTCCTACTGT TCAAGTACAGGGGCTGGTCCGCAAAGGGAAGAAAAGCAAAAGACGAAAATGGCTAAA TTCGTGATCCGCCAGCCACTGCCGCCGACTGCAGTGACATACTGCGGCTGATCAAGG AGCTGGCTAAATATGAATACATCGGAAGAACAAGTAATCTTAACTGAAAAAGATCTGCT AGAAGATGGTTTTGGAGAGCACCCCTTTTACCACCTGCCTGGTTGCAGAAGTGCCGAAA GAGCACTGGACTCCGGAAGGACACAGCATTGTTGGTTTTGCCATGTACTATTTTACCT ATGACCCGTGGATTGGCAAGTTATTGTATCTTGAGGACTTCTTCGTGATGAGTGATTA TAGAGGCTTTGGCATAGGATCAGAAATTCGAAGAATCTAAGCCAGGTTGCAATGAGG TGTCGCTGCAGCAGCATGCACCTTCTGGTAGCAGAATGGAATGAACCATCCATCAACT TCTATAAAGAAGAGGTGCTTCTGATCTGTCCAGTGAAGAGGGTTGGAGACTGTTCAA GATCGACAAGGAGTACTTGCTAAAAATGGCAACAGAGGAGTGAGGAGTGCTGCTGTAG ATGACAACCTCCATTCTATTTTAGAATAAATCCCAACTTCTCTTGCTTTCTATGCTG TTTGTAGTGAAATAATAGAATGAGCACCCATTCCAAAGCTTTATTACCAGTGGCGTTG TTGCATGTTTGAAATGAGGTCTGTTTAAAGTGGCAATCTCAGATGCAGTTTGGAGAGT CAGATCTTCTCCTTGAATATCTTTCGATAAAACAACAAGGTGGTGTGATCTTAATATA TTTGAAAAAACTTCATTCTCGTGAGTCATTTAAATGTGTACAATGTACACACTGGTA CTTAGAGTTTCTGTTGATTCTTTTAAATAAACTACTCTTTGATTAAAAA AAAAAAAAAAAAAAAA		
	ORF Start: ATG at 166		ORF Stop: TGA at 679
	SEQ ID NO: 442	171 aa	MW at 20023.8kD
NOV59a,	MAKFVIRPATAADCSILRLIKELAKYEYMEEQVILTEKDLEDFGEHPFYHCLVAE		

CG98102-01 Protein Sequence	VPKEHWTPEGHSIVGFAMYFTYDPWIGKLLYLEDDFFVMSDYRFGIGSEILKNLSQV AMRRCSSMHFLVAEWNPSINFYKRRGASDLSSSEGWRLFIDKEYLLKMATEE		
	SEQ ID NO: 443	1052 bp	
NOV59b, CG98102-03 DNA Sequence	CGGCCGCGTCGACCGCGGGCTGACTGGTGTTCCTCCGTCCTCGCCGAGGTTCCCTTGG GTCATGGTGCCAGCCTGACTGAGAAGAGGACGCTCCCGGGAGACGAATGAGGAACAC CTCCTCCTACTGTTCAAGTACAGGGGCTGGTCCGCAAAGGGAAGAAAAGCAAAGAC GAAAATGGCTAAATTCGTGATCCGCCAGCCACTGCCGCCGACTGCAGTGACATACTG CGGCTGATCAAGGAGCTGGCTAAATATGAATACATGGAAGAACAAGTAATCTTAAGT AAAAAGATCTGCTAGAAGATGGTTTTGGAGAGCACCCCTTTTACCAGTGCCTGGTTGC AGAAGTGCCGAAAGAGCACTGGACTCCGGAAGGACACAGCATTGTTGGTTTTGCCATG TACTATTTTACCTATGACCCGTGGATTGGCAAGTTATTGTATCTTGAGGACTTTTTTCG TGATGAGTGATTATAGAGGCTTTGGCATAGGATCAGAAATCTGAAGAATCTAAGCCA GGTTGCAATGAGGTGTCGCTGCAGCAGCATGCACCTCTTGGTAGCAGAATGGAATGAA CCATCCATCAACTTCTATAAAGAAGAGGTGCTTCTGATCTGTCCAGTGAAGAGGGTT GGAGACTGTTCAAGATCGACAAGGAGTACTTGCTAAAAATGGCAACAGAGGAGTGAGG AGTGCTGCTGTAGATGACAACCTCCATTCTATTTTGAATAAAATCCCAACTTCTCTT GCTTTCTATGCTGTTTGTAGTGAAAATAATAGAATGAGCACCCATTCCAAAGCTTTATT ACCAGTGGCGTTGTTGCATGTTTGAAATGAGGTCTGTTTAAAGTGGCAATCTCAGATG CAGTTTGGAGAGTCAGATCTTTCTCCTTGAATATCTTTTCGATAAACAACAGGTGGTG TGATCTTAATATATTTGAAAAAACTTCATTCTCGTGAGTCAATTAATGTGTACAAT GTACACACTGGTACTTAGAGTTTCTGTTTGATTCTTTTTTAATAAACTACTCTTTGAT TTAAAAAA		
	ORF Start: ATG at 179		ORF Stop: TGA at 692
	SEQ ID NO: 444	171 aa	MW at 20023.8kD
NOV59b, CG98102-03 Protein Sequence	MAKFVIRPATAADCSIDLRLIKELAKYEYMEEQVILTEKDLEDGFGHEHPFYHCLVAE VPKEHWTPEGHSIVGFAMYFTYDPWIGKLLYLEDDFFVMSDYRFGIGSEILKNLSQV AMRRCSSMHFLVAEWNPSINFYKRRGASDLSSSEGWRLFIDKEYLLKMATEE		
	SEQ ID NO: 445	665 bp	
NOV59c, CG98102-02 DNA Sequence	ACCTCCTCCTACTGTTCAAGTACAGGGGCTGGTCCGCAAAGGGAAGAAAAGCAAAG ACGAAAATGGCTAAATTCGTGATCCGCCAGCCACTGCCGCCGACTGCAGTGACATAC TGCGGCTGATCAAGGAGCTGGCTAAATATGAATACATGGAAGAACAAGTAATCTTAAC TGAAAAAGATCTGCTAGAAGATGGTTTTGGAGAGCACCCCTTTTACCAGTGCCTGGTT GCAGAAGTGCCGAAAGAGCACTGGACTCCGGAAGGACACAGCATTGTTGGTTTTGCCA TGTAATATTTTACCTATGACCCGTGGATTGGCAAGTTATTGTATCTTGAGGACTTCTT CGTGATGAGTGATTATAGAGGCTTTGGCATAGGATCAGAAATCTGAAGAATCTAAGC CAGGTTGCAATGAGGTGTCGCTGCAGCAGCATGCACCTCTTGGTAGCAGAATGGAATG AACCATCCATCAACTTCTATAAAGAAGAGGTGCTTCTGATCTGTCCAGTGAAGAGGG TTGGAGACTGTTCAAGATCGACAAGGAGTACTTGCTAAAAATGGCAACAGAGGAGTGA GGAGTGCTGCTGTAGATGACAACCTCCATTCTATTTTGAATAAAATCCCAACTTCTC TTGCTTTCTATGCTGTTTGTAGTGAAA		
	ORF Start: ATG at 65		ORF Stop: TGA at 578
	SEQ ID NO: 446	171 aa	MW at 20023.8kD
NOV59c, CG98102-02 Protein Sequence	MAKFVIRPATAADCSIDLRLIKELAKYEYMEEQVILTEKDLEDGFGHEHPFYHCLVAE VPKEHWTPEGHSIVGFAMYFTYDPWIGKLLYLEDDFFVMSDYRFGIGSEILKNLSQV AMRRCSSMHFLVAEWNPSINFYKRRGASDLSSSEGWRLFIDKEYLLKMATEE		
	SEQ ID NO: 447	596 bp	
NOV59d, CG98102-04 DNA Sequence	ACCTCCTCCTACTGTTCAAGTACAGGGGCTGGTCCGCAAAGGGAAGAAAAGCAAAG ACGAAAATGGCTAAATATGAATACATGGAAGAACAAGTAATCTTAAGTAAAAAGATC TGCTAGAAGATGGTTTTGGAGAGCACCCCTTTTACCAGTGCCTGGTTGCAGAAGTGCC GAAAGAGCACTGGACTCCGGAAGGACACAGCATTGTTGGTTTTGCCATGTACTATTTT ACCTATGACCCGTGGATTGGCAAGTTATTGTATCTTGAGGACTTCTTCGTGATGAGTG ATTATAGAGGCTTTGGCATAGGATCAGAAATCTGAAGAATCTAAGCCAGGTTGCAAT GAGGTGCTGCTGCAGCAGCATGCACCTCTTGGTAGCAGAATGGAATGAACATCCATC		

	AACTTCTATAAAAGAAGAGGTGCTTCTGATCTGTCCAGTGAAGAGGGTTGGAGACTGT TCAAGATCGACAAGGAGTACTTGCTAAAAATGGCAACAGAGGAGTGAGGAGTGCTGCT GTAGATGACAACCTCCATTCTATTTTAGAATAAAATCCCAACTTCTCTTGCTTTCTAT GCTGTTTGTAGTGA		
	ORF Start: ATG at 65		ORF Stop: TGA at 509
	SEQ ID NO: 448	148 aa	MW at 17497.8kD
NOV59d, CG98102-04 Protein Sequence	MAKYEYMEEQVILTEKDLLEDGFGEHPFYHCLVAEVPKEHWTPEGHSIVGFAMYFTY DPWIGKLLYLEDDFFVMSDYRGFGIGSEILKNLSQVAMRRCSSMHFLVAEWNEPSINF YKRRGASDLSSSEGWRLFKIDKEYLLKMATEE		
	SEQ ID NO: 449	1157 bp	
NOV59e, CG98102-05 DNA Sequence	CTGGTGTATTATCCGTCACCTCGCCGAGGTTCCCTGGGTTCATGGTGCCAGCCTGACTGAG AAGAGGACGCTCCCGGAGACGAATGAGGAACCACTCCTCCTACTGTTCAAGTACAG GGGCTTGGTCCGCAAAGGGAAGAAAAGCAAAAGACGAAAATGGCTAAATTCGTGATCC GCCAGCCACTGCCGCCGACTGCAGTGACATACTGCGGCTGATCAAGGAGCTGGCTAA ATATGAATACATGGAAGAACAAGTAATCTTAACGAAAAAGATCTGCTAGAAGATGGT TTTGGAGAGCACCCCTTTTACCCTGCCTGGTTGCAGAAAGTCCGAAAGAGCACTGGA CTCCGGAAGGTTACAGTCTCTAGCTTCGCCATGTACATGGCCCTTCCGTGTACATGGA TGGGCGGGGAGGTAACATAAAGATCCTTTACACAATAAAGTAGATGATCATGATAAAT GAGGACACAGCATTGTTGGTTTTGCCATGTACTATTTTACCCTATGACCCGTGGATTGG CAAGTTATTGTATCTTGAGGACTTCTTCGTGATGAGTGATTATAGAGGCTTTGGCATA GGATCAGAAATTCTGAAGAATCTAAGCCAGGTTGCAATGAGGTGTCGTCGAGCAGCA TGCACCTCTTGGTAGCAGAATGGAATGAACCATCCATCAACTTCTATAAAGAAAGAGG TGCTTCTGATCTGTCCAGTGAAGAGGGTTGGAGACTGTTCAAGATCGACAAGGAGTAC TTGCTAAAAATGGCAACAGAGGAGTGAGGAGTGCTGCTGTAGATGACAACCTCCATTCT TATTTTAGAATAAAATCCCAACTTCTCTTGCTTTCTATGCTGTTTGTAGTGAAATAAT AGAATGAGCACCCATTCCAAAGCTTTATTACCAGTGGCGTTGTTGCATGTTTGAAATG AGGTCTGTTTAAAGTGGCAATCTCAGATGCAGTTTGGAGAGTCAGATCTTCTCCTTG AATATCTTTGATAAACAACAAGGTGGTGTGATCTTAATATATTGAAAAAACTTCA TTCTCGTGAGTCATTTAAATGTGTACAATGTACACACTGGTACTTAGAGTTTCTGTTT GATTCTTTTTTAATAAACTACTCTTTGATTAAAAA		
	ORF Start: ATG at 491		ORF Stop: TGA at 779
	SEQ ID NO: 450	96 aa	MW at 11464.0kD
NOV59e, CG98102-05 Protein Sequence	MYFTYDPWIGKLLYLEDDFFVMSDYRGFGIGSEILKNLSQVAMRRCSSMHFLVAEWN EPSINFYKRRGASDLSSSEGWRLFKIDKEYLLKMATEE		
	SEQ ID NO: 451	1107 bp	
NOV59f, CG98102-06 DNA Sequence	TGGAATTCGGCCATACTGGCGGTAGCGAGCTCTTAGTCGCGGGCCGACTGGTGT ATCCGTCACCTGCGGAGGTTCCCTGGGTTCATGGTGCCAGCCTGACTGAGAAGAGGACG CTCCCGGGAGACGAATGAGTGAACCACTCCTCCTACTGTTCAAGTACAGGGGCTGG TCCGCAAAGGGAAGAAAAGCAAAAGACGAAAATGGCTAAATTCGTGATCCGCCAGCC ACTGCCGCCGACTGCAGTGACATACTGCGGCTGATCAAGGAGCTGGCTAAATATGAAT ACATGGAAGAACAAGTAATCTTAACGAAAAAGATCTGCTAGAAGATGGTTTTGGAGA GCACCCCTTTTACCCTGCCTGGTTGCAGAAAGTCCGAAAGAGCACTGGACTCCGGAA GGTAACCCCTCGCCCTTGTCCAGGTAAGCCATGTAGTAGTTTACCTATACCCGTGTT ATGTAAGCAAGTTATGGTGTCTTGAGGACTTCTTCGTGATGAGTGATTACTCGAGGCT TTGGCATAGGATCAGAAATCTGAAGAATCTAAGCCAGGTTGCAATGAGGTGTCGCTG CCAGCAGCATGCACCTTTGGGTAGCAGAATGGAATGAACCATCCATCAACTTCTATA AAAGAAGAGGTGCTTCTGATCTGTCCAGTGAAGAGGGTTGGAGATGTTTCAGATCGCA GGAGTACTGCTAAAAATGGCAACAGGGAGTACGAGACTGTGCTGATAGATGACAACCT CCATTCTATTTTAGAATAAAATCCCAACTTCTCTTGCTTTCTATGCTGTTTGTAGTGA AATAATAGAATGAGCACCCATTCCAAAGCTTTATTACCAGTGGCGTTGTTGCATGTTT GAAATGAGGTCTGTTTAAAGTGGCAATCTCAGATGCAGTTTGGAGAGTCAGATCTTTC TCCTTGAATATCTTTGATAAACAACAAGGTGGTGTGATCTTAATATATTGAAAAAA ACTTCAATTCTCGTGAGTCATTTAAATGTGTACAATGTACACACTGGTACTTAGAGTTT CTGTTTGATTCTTTTTTAATAAACTACTCTTTGATTAAAAA		

	AAAAA		
	ORF Start: ATG at 131		ORF Stop: TAA at 707
	SEQ ID NO: 452	192 aa	MW at 22209.9kD
NOV59f, CG98102-06 Protein Sequence	MSEPPPPPTVQVQGGPGPQREEKQKTKMAKFVIRPATAADCSIDLRLIKELAKYEYMEEQ VILTEKDILLEDGFGEHPFYHCLVAEVPKEHWTPEGNPSPLSRVSHVVVLYPCYVSKL WCLEDDFFVMSDYSRLWHRIRNSEESKPGCNEVSLPAACTSWVAEWNPEPSINFYKRRGA SDLSSEEGWRCSDRKEYC		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 59B.

Table 59B. Comparison of NOV59a against NOV59b through NOV59f.		
Protein Sequence	NOV59a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV59b	1..171 1..171	171/171 (100%) 171/171 (100%)
NOV59c	1..171 1..171	171/171 (100%) 171/171 (100%)
NOV59d	24..171 1..148	147/148 (99%) 148/148 (99%)
NOV59e	76..171 1..96	96/96 (100%) 96/96 (100%)
NOV59f	1..155 26..184	115/163 (70%) 124/163 (75%)

- 5 Further analysis of the NOV59a protein yielded the following properties shown in Table 59C.

Table 59C. Protein Sequence Properties NOV59a	
PSort analysis:	0.6400 probability located in microbody (peroxisome); 0.6153 probability located in mitochondrial matrix space; 0.3177 probability located in mitochondrial inner membrane; 0.3177 probability located in mitochondrial intermembrane space
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV59a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 59D.

Table 59D. Geneseq Results for NOV59a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV59a Residues/	Identities/ Similarities for	Expect Value

		<b>Match Residues</b>	<b>the Matched Region</b>	
ABB57094	Mouse ischaemic condition related protein sequence SEQ ID NO:207 - <i>Mus musculus</i> , 171 aa. [WO200188188-A2, 22-NOV-2001]	1..171 1..171	165/171 (96%) 168/171 (97%)	1e-96
AAU30048	Novel human secreted protein #539 - <i>Homo sapiens</i> , 218 aa. [WO200179449-A2, 25-OCT-2001]	1..158 35..195	146/161 (90%) 151/161 (93%)	9e-81
AAB82049	Human spermidine/spermine acetyl transferase protein isomer - <i>Homo sapiens</i> , 192 aa. [CN1278003-A, 27-DEC-2000]	1..155 26..184	115/163 (70%) 124/163 (75%)	4e-56
AAB44145	Human cancer associated protein sequence SEQ ID NO:1590 - <i>Homo sapiens</i> , 92 aa. [WO200055350-A1, 21-SEP-2000]	42..127 1..86	85/86 (98%) 85/86 (98%)	3e-48
AAW58394	Human spermidine/spermine N1-acetyltransferase - <i>Homo sapiens</i> , 170 aa. [WO9818938-A1, 07-MAY-1998]	1..168 1..168	78/168 (46%) 109/168 (64%)	9e-41

In a BLAST search of public sequence databases, the NOV59a protein was found to have homology to the proteins shown in the BLASTP data in Table 59E.

**Table 59E. Public BLASTP Results for NOV59a**

<b>Protein Accession Number</b>	<b>Protein/Organism/Length</b>	<b>NOV59a Residues/ Match Residues</b>	<b>Identities/ Similarities for the Matched Portion</b>	<b>Expect Value</b>
P21673	Diamine acetyltransferase (EC 2.3.1.57) (Spermidine/spermine N(1)- acetyltransferase) (SSAT) (Putrescine acetyltransferase) - <i>Homo sapiens</i> (Human), 171 aa.	1..171 1..171	171/171 (100%) 171/171 (100%)	3e-99
JH0783	diamine N-acetyltransferase (EC 2.3.1.57) - human, 171 aa.	1..171 1..171	170/171 (99%) 171/171 (99%)	1e-98
P49431	Spermidine/spermine N(1)-acetyltransferase (EC 2.3.1.57) (Diamine acetyltransferase) (SSAT) (Putrescine acetyltransferase) - <i>Mus saxicola</i> (Spiny mouse), 171 aa.	1..171 1..171	166/171 (97%) 169/171 (98%)	7e-97

Q28999	Diamine acetyltransferase (EC 2.3.1.57) (Spermidine/spermine N(1)- acetyltransferase) (SSAT) (Putrescine acetyltransferase) - Sus scrofa (Pig), 171 aa.	1..171 1..171	168/171 (98%) 169/171 (98%)	1e-96
Q9JHW6	Spermidine/spermine N1- acetyltransferase - Cricetulus griseus (Chinese hamster), 171 aa.	1..171 1..171	164/171 (95%) 169/171 (97%)	2e-96

PFam analysis predicts that the NOV59a protein contains the domains shown in the Table 59F.

Table 59F. Domain Analysis of NOV59a			
Pfam Domain	NOV59a Match Region	Identities/ Similarities for the Matched Region	Expect Value
Acetyltransf	63..146	23/85 (27%) 59/85 (69%)	1.6e-16

**Example B: Sequencing Methodology and Identification of NOVX Clones**

1. **GeneCalling™ Technology:** This is a proprietary method of performing differential gene expression profiling between two or more samples developed at CuraGen and described by Shimkets, et al., "Gene expression analysis by transcript profiling coupled to a gene database query" Nature Biotechnology 17:198-803 (1999). cDNA was derived from various human samples representing multiple tissue types, normal and diseased states, physiological states, and developmental states from different donors. Samples were obtained as whole tissue, primary cells or tissue cultured primary cells or cell lines. Cells and cell lines may have been treated with biological or chemical agents that regulate gene expression, for example, growth factors, chemokines or steroids. The cDNA thus derived was then digested with up to as many as 120 pairs of restriction enzymes and pairs of linker-adaptors specific for each pair of restriction enzymes were ligated to the appropriate end. The restriction digestion generates a mixture of unique cDNA gene fragments. Limited PCR amplification is performed with primers homologous to the linker adapter sequence where one primer is biotinylated and the other is fluorescently labeled. The doubly labeled material is isolated and the fluorescently labeled single strand is resolved by capillary gel electrophoresis. A computer algorithm compares the electropherograms from an experimental and control group for each of the restriction digestions. This and additional sequence-derived information is used to predict the identity of each differentially expressed gene fragment using a variety of genetic databases. The identity of the gene fragment is confirmed by additional, gene-specific competitive PCR or by isolation and sequencing of the gene fragment.

2. **SeqCalling™ Technology:** cDNA was derived from various human samples representing multiple tissue types, normal and diseased states, physiological states, and developmental states from different donors. Samples were obtained as whole tissue, primary cells or tissue cultured primary cells or cell lines. Cells and cell lines may have been treated with biological or chemical agents that regulate gene expression, for example, growth factors, chemokines or steroids. The cDNA thus derived was then sequenced using CuraGen's proprietary SeqCalling technology. Sequence traces were evaluated manually and edited for corrections if appropriate. cDNA sequences from all samples were assembled together, sometimes including public human sequences, using bioinformatic programs to produce a consensus sequence for each assembly. Each assembly is included in CuraGen Corporation's database. Sequences were included as components for assembly

when the extent of identity with another component was at least 95% over 50 bp. Each assembly represents a gene or portion thereof and includes information on variants, such as splice forms single nucleotide polymorphisms (SNPs), insertions, deletions and other sequence variations.

5           **3. PathCalling™ Technology:** The NOVX nucleic acid sequences are derived by laboratory screening of cDNA library by the two-hybrid approach. cDNA fragments covering either the full length of the DNA sequence, or part of the sequence, or both, are sequenced. In silico prediction was based on sequences available in CuraGen Corporation's proprietary sequence databases or in the public human sequence databases,  
10 and provided either the full length DNA sequence, or some portion thereof.

The laboratory screening was performed using the methods summarized below:

cDNA libraries were derived from various human samples representing multiple tissue types, normal and diseased states, physiological states, and developmental states from different donors. Samples were obtained as whole tissue, primary cells or tissue  
15 cultured primary cells or cell lines. Cells and cell lines may have been treated with biological or chemical agents that regulate gene expression, for example, growth factors, chemokines or steroids. The cDNA thus derived was then directionally cloned into the appropriate two-hybrid vector (Gal4-activation domain (Gal4-AD) fusion). Such cDNA libraries as well as commercially available cDNA libraries from Clontech (Palo Alto, CA)  
20 were then transferred from E.coli into a CuraGen Corporation proprietary yeast strain (disclosed in U. S. Patents 6,057,101 and 6,083,693, incorporated herein by reference in their entireties).

Gal4-binding domain (Gal4-BD) fusions of a CuraGen Corporation proprietary library of human sequences was used to screen multiple Gal4-AD fusion cDNA libraries  
25 resulting in the selection of yeast hybrid diploids in each of which the Gal4-AD fusion contains an individual cDNA. Each sample was amplified using the polymerase chain reaction (PCR) using non-specific primers at the cDNA insert boundaries. Such PCR product was sequenced; sequence traces were evaluated manually and edited for corrections if appropriate. cDNA sequences from all samples were assembled together,  
30 sometimes including public human sequences, using bioinformatic programs to produce a consensus sequence for each assembly. Each assembly is included in CuraGen Corporation's database. Sequences were included as components for assembly when the extent of identity with another component was at least 95% over 50 bp. Each assembly

represents a gene or portion thereof and includes information on variants, such as splice forms single nucleotide polymorphisms (SNPs), insertions, deletions and other sequence variations.

Physical clone: the cDNA fragment derived by the screening procedure, covering  
5 the entire open reading frame is, as a recombinant DNA, cloned into pACT2 plasmid (Clontech) used to make the cDNA library. The recombinant plasmid is inserted into the host and selected by the yeast hybrid diploid generated during the screening procedure by the mating of both CuraGen Corporation proprietary yeast strains N106' and YULH (U. S. Patents 6,057,101 and 6,083,693).

10           4.     **RACE:** Techniques based on the polymerase chain reaction such as rapid amplification of cDNA ends (RACE), were used to isolate or complete the predicted sequence of the cDNA of the invention. Usually multiple clones were sequenced from one or more human samples to derive the sequences for fragments. Various human tissue samples from different donors were used for the RACE reaction. The sequences derived  
15 from these procedures were included in the SeqCalling Assembly process described in preceding paragraphs.

5.     **Exon Linking:** The NOVX target sequences identified in the present invention were subjected to the exon linking process to confirm the sequence. PCR primers were designed by starting at the most upstream sequence available, for the forward  
20 primer, and at the most downstream sequence available for the reverse primer. In each case, the sequence was examined, walking inward from the respective termini toward the coding sequence, until a suitable sequence that is either unique or highly selective was encountered, or, in the case of the reverse primer, until the stop codon was reached. Such primers were designed based on in silico predictions for the full length cDNA, part (one or  
25 more exons) of the DNA or protein sequence of the target sequence, or by translated homology of the predicted exons to closely related human sequences from other species. These primers were then employed in PCR amplification based on the following pool of human cDNAs: adrenal gland, bone marrow, brain - amygdala, brain - cerebellum, brain - hippocampus, brain - substantia nigra, brain - thalamus, brain -whole, fetal brain, fetal  
30 kidney, fetal liver, fetal lung, heart, kidney, lymphoma - Raji, mammary gland, pancreas, pituitary gland, placenta, prostate, salivary gland, skeletal muscle, small intestine, spinal cord, spleen, stomach, testis, thyroid, trachea, uterus. Usually the resulting amplicons were gel purified, cloned and sequenced to high redundancy. The PCR product derived from

exon linking was cloned into the pCR2.1 vector from Invitrogen. The resulting bacterial clone has an insert covering the entire open reading frame cloned into the pCR2.1 vector. The resulting sequences from all clones were assembled with themselves, with other fragments in CuraGen Corporation's database and with public ESTs. Fragments and ESTs were included as components for an assembly when the extent of their identity with another component of the assembly was at least 95% over 50 bp. In addition, sequence traces were evaluated manually and edited for corrections if appropriate. These procedures provide the sequence reported herein.

6. **Physical Clone:** Exons were predicted by homology and the intron/exon boundaries were determined using standard genetic rules. Exons were further selected and refined by means of similarity determination using multiple BLAST (for example, tBlastN, BlastX, and BlastN) searches, and, in some instances, GeneScan and Grail. Expressed sequences from both public and proprietary databases were also added when available to further define and complete the gene sequence. The DNA sequence was then manually corrected for apparent inconsistencies thereby obtaining the sequences encoding the full-length protein.

The PCR product derived by exon linking, covering the entire open reading frame, was cloned into the pCR2.1 vector from Invitrogen to provide clones used for expression and screening purposes.

#### **Example C: Quantitative expression analysis of clones in various cells and tissues**

The quantitative expression of various clones was assessed using microtiter plates containing RNA samples from a variety of normal and pathology-derived cells, cell lines and tissues using real time quantitative PCR (RTQ PCR). RTQ PCR was performed on an Applied Biosystems ABI PRISM® 7700 or an ABI PRISM® 7900 HT Sequence Detection System. Various collections of samples are assembled on the plates, and referred to as Panel 1 (containing normal tissues and cancer cell lines), Panel 2 (containing samples derived from tissues from normal and cancer sources), Panel 3 (containing cancer cell lines), Panel 4 (containing cells and cell lines from normal tissues and cells related to inflammatory conditions), Panel 5D/5I (containing human tissues and cell lines with an emphasis on metabolic diseases), AI\_comprehensive\_panel (containing normal tissue and samples from autoimmune/autoinflammatory diseases), Panel CNSD.01 (containing

samples from normal and diseased brains) and CNS\_neurodegeneration\_panel (containing samples from normal and Alzheimer's diseased brains).

RNA integrity from all samples is controlled for quality by visual assessment of agarose gel electropherograms using 28S and 18S ribosomal RNA staining intensity ratio as a guide (2:1 to 2.5:1 28s:18s) and the absence of low molecular weight RNAs that would be indicative of degradation products. Samples are controlled against genomic DNA contamination by RTQ PCR reactions run in the absence of reverse transcriptase using probe and primer sets designed to amplify across the span of a single exon.

First, the RNA samples were normalized to reference nucleic acids such as constitutively expressed genes (for example,  $\beta$ -actin and GAPDH). Normalized RNA (5  $\mu$ l) was converted to cDNA and analyzed by RTQ-PCR using One Step RT-PCR Master Mix Reagents (Applied Biosystems; Catalog No. 4309169) and gene-specific primers according to the manufacturer's instructions.

In other cases, non-normalized RNA samples were converted to single strand cDNA (sscDNA) using Superscript II (Invitrogen Corporation; Catalog No. 18064-147) and random hexamers according to the manufacturer's instructions. Reactions containing up to 10  $\mu$ g of total RNA were performed in a volume of 20  $\mu$ l and incubated for 60 minutes at 42 °C. This reaction can be scaled up to 50  $\mu$ g of total RNA in a final volume of 100  $\mu$ l. sscDNA samples are then normalized to reference nucleic acids as described previously, using 1X TaqMan® Universal Master mix (Applied Biosystems; catalog No. 4324020), following the manufacturer's instructions.

Probes and primers were designed for each assay according to Applied Biosystems Primer Express Software package (version I for Apple Computer's Macintosh Power PC) or a similar algorithm using the target sequence as input. Default settings were used for reaction conditions and the following parameters were set before selecting primers: primer concentration = 250 nM, primer melting temperature ( $T_m$ ) range = 58 °-60 °C, primer optimal  $T_m$  = 59 °C, maximum primer difference = 2 °C, probe does not have 5'G, probe  $T_m$  must be 10 °C greater than primer  $T_m$ , amplicon size 75bp to 100bp. The probes and primers selected (see below) were synthesized by Synthegen (Houston, TX, USA). Probes were double purified by HPLC to remove uncoupled dye and evaluated by mass spectroscopy to verify coupling of reporter and quencher dyes to the 5' and 3' ends of the probe, respectively. Their final concentrations were: forward and reverse primers, 900nM each, and probe, 200nM.

PCR conditions: When working with RNA samples, normalized RNA from each tissue and each cell line was spotted in each well of either a 96 well or a 384-well PCR plate (Applied Biosystems). PCR cocktails included either a single gene specific probe and primers set, or two multiplexed probe and primers sets (a set specific for the target clone and another gene-specific set multiplexed with the target probe). PCR reactions were set up using TaqMan® One-Step RT-PCR Master Mix (Applied Biosystems, Catalog No. 4313803) following manufacturer's instructions. Reverse transcription was performed at 48°C for 30 minutes followed by amplification/PCR cycles as follows: 95°C 10 min, then 40 cycles of 95 °C for 15 seconds, 60 °C for 1 minute. Results were recorded as CT values (cycle at which a given sample crosses a threshold level of fluorescence) using a log scale, with the difference in RNA concentration between a given sample and the sample with the lowest CT value being represented as 2 to the power of delta CT. The percent relative expression is then obtained by taking the reciprocal of this RNA difference and multiplying by 100.

When working with sscDNA samples, normalized sscDNA was used as described previously for RNA samples. PCR reactions containing one or two sets of probe and primers were set up as described previously, using 1X TaqMan® Universal Master mix (Applied Biosystems; catalog No. 4324020), following the manufacturer's instructions. PCR amplification was performed as follows: 95 °C 10 min, then 40 cycles of 95 °C for 15 seconds, 60 °C for 1 minute. Results were analyzed and processed as described previously.

#### **Panels 1, 1.1, 1.2, and 1.3D**

The plates for Panels 1, 1.1, 1.2 and 1.3D include 2 control wells (genomic DNA control and chemistry control) and 94 wells containing cDNA from various samples. The samples in these panels are broken into 2 classes: samples derived from cultured cell lines and samples derived from primary normal tissues. The cell lines are derived from cancers of the following types: lung cancer, breast cancer, melanoma, colon cancer, prostate cancer, CNS cancer, squamous cell carcinoma, ovarian cancer, liver cancer, renal cancer, gastric cancer and pancreatic cancer. Cell lines used in these panels are widely available through the American Type Culture Collection (ATCC), a repository for cultured cell lines, and were cultured using the conditions recommended by the ATCC. The normal tissues found on these panels are comprised of samples derived from all major organ systems from single adult individuals or fetuses. These samples are derived from the following organs: adult

skeletal muscle, fetal skeletal muscle, adult heart, fetal heart, adult kidney, fetal kidney, adult liver, fetal liver, adult lung, fetal lung, various regions of the brain, the spleen, bone marrow, lymph node, pancreas, salivary gland, pituitary gland, adrenal gland, spinal cord, thymus, stomach, small intestine, colon, bladder, trachea, breast, ovary, uterus, placenta, prostate, testis and adipose.

In the results for Panels 1, 1.1, 1.2 and 1.3D, the following abbreviations are used:

ca. = carcinoma,  
 \* = established from metastasis,  
 met = metastasis,  
 10 s cell var = small cell variant,  
 non-s = non-sm = non-small,  
 squam = squamous,  
 pl. eff = pl effusion = pleural effusion,  
 glio = glioma,  
 15 astro = astrocytoma, and  
 neuro = neuroblastoma.

#### **General\_screening\_panel\_v1.4, v1.5, v1.6 and 1.7**

The plates for Panels 1.4, 1.5, 1.6 and 1.7 include 2 control wells (genomic DNA control and chemistry control) and 94 wells containing cDNA from various samples. The samples in Panels 1.4, 1.5, 1.6 and 1.7 are broken into 2 classes: samples derived from cultured cell lines and samples derived from primary normal tissues. The cell lines are derived from cancers of the following types: lung cancer, breast cancer, melanoma, colon cancer, prostate cancer, CNS cancer, squamous cell carcinoma, ovarian cancer, liver cancer, renal cancer, gastric cancer and pancreatic cancer. Cell lines used in Panels 1.4, 1.5, and 1.6 are widely available through the American Type Culture Collection (ATCC), a repository for cultured cell lines, and were cultured using the conditions recommended by the ATCC. The normal tissues found on Panels 1.4, 1.5, 1.6, 1.7 are comprised of pools of samples derived from all major organ systems from 2 to 5 different adult individuals or fetuses. These samples are derived from the following organs: adult skeletal muscle, fetal skeletal muscle, adult heart, fetal heart, adult kidney, fetal kidney, adult liver, fetal liver, adult lung, fetal lung, various regions of the brain, the spleen, bone marrow, lymph node, pancreas, salivary gland, pituitary gland, adrenal gland, spinal cord, thymus, stomach, small intestine, colon,

bladder, trachea, breast, ovary, uterus, placenta, prostate, testis and adipose. Abbreviations are as described for Panels 1, 1.1, 1.2, and 1.3D.

#### **Panels 2D, 2.2, 2.3 and 2.4**

5           The plates for Panels 2D, 2.2, 2.3 and 2.4 generally include two control wells and 94 test samples composed of RNA or cDNA isolated from human tissue procured by surgeons working in close cooperation with the National Cancer Institute's Cooperative Human Tissue Network (CHTN) or the National Disease Research Initiative (NDRI) or from Ardaïs or Clinomics. The tissues are derived from human malignancies and in cases  
10       where indicated many malignant tissues have "matched margins" obtained from noncancerous tissue just adjacent to the tumor. These are termed normal adjacent tissues and are denoted "NAT" in the results below. The tumor tissue and the "matched margins" are evaluated by two independent pathologists (the surgical pathologists and again by a pathologist at NDRI/ CHTN/Ardaïs/Clinomics). Unmatched RNA samples from tissues  
15       without malignancy (normal tissues) were also obtained from Ardaïs or Clinomics. This analysis provides a gross histopathological assessment of tumor differentiation grade. Moreover, most samples include the original surgical pathology report that provides information regarding the clinical stage of the patient. These matched margins are taken from the tissue surrounding (*i.e.* immediately proximal) to the zone of surgery (designated  
20       "NAT", for normal adjacent tissue, in Table RR). In addition, RNA and cDNA samples were obtained from various human tissues derived from autopsies performed on elderly people or sudden death victims (accidents, *etc.*). These tissues were ascertained to be free of disease and were purchased from various commercial sources such as Clontech (Palo Alto, CA), Research Genetics, and Invitrogen. General oncology screening panel\_v\_2.4 is  
25       an updated version of Panel 2D.

#### **HASS Panel v 1.0**

          The HASS panel v 1.0 plates are comprised of 93 cDNA samples and two controls. Specifically, 81 of these samples are derived from cultured human cancer cell lines that had been subjected to serum starvation, acidosis and anoxia for different time periods as well as  
30       controls for these treatments, 3 samples of human primary cells, 9 samples of malignant brain cancer (4 medulloblastomas and 5 glioblastomas) and 2 controls. The human cancer cell lines are obtained from ATCC (American Type Culture Collection) and fall into the following tissue groups: breast cancer, prostate cancer, bladder carcinomas, pancreatic

cancers and CNS cancer cell lines. These cancer cells are all cultured under standard recommended conditions. The treatments used (serum starvation, acidosis and anoxia) have been previously published in the scientific literature. The primary human cells were obtained from Clonetics (Walkersville, MD) and were grown in the media and conditions recommended by Clonetics. The malignant brain cancer samples are obtained as part of a collaboration (Henry Ford Cancer Center) and are evaluated by a pathologist prior to CuraGen receiving the samples. RNA was prepared from these samples using the standard procedures. The genomic and chemistry control wells have been described previously.

#### 10           **ARDAIS Panel v 1.0**

The plates for ARDAIS panel v 1.0 generally include 2 control wells and 22 test samples composed of RNA isolated from human tissue procured by surgeons working in close cooperation with Ardaais Corporation. The tissues are derived from human lung malignancies (lung adenocarcinoma or lung squamous cell carcinoma) and in cases where indicated many malignant samples have “matched margins” obtained from noncancerous lung tissue just adjacent to the tumor. These matched margins are taken from the tissue surrounding (*i.e.* immediately proximal) to the zone of surgery (designated “NAT”, for normal adjacent tissue) in the results below. The tumor tissue and the “matched margins” are evaluated by independent pathologists (the surgical pathologists and again by a pathologist at Ardaais). Unmatched malignant and non-malignant RNA samples from lungs were also obtained from Ardaais. Additional information from Ardaais provides a gross histopathological assessment of tumor differentiation grade and stage. Moreover, most samples include the original surgical pathology report that provides information regarding the clinical state of the patient.

#### 25           **Panel 3D, 3.1 and 3.2**

The plates of Panel 3D, 3.1, and 3.2 are comprised of 94 cDNA samples and two control samples. Specifically, 92 of these samples are derived from cultured human cancer cell lines, 2 samples of human primary cerebellar tissue and 2 controls. The human cell lines are generally obtained from ATCC (American Type Culture Collection), NCI or the German tumor cell bank and fall into the following tissue groups: Squamous cell carcinoma of the tongue, breast cancer, prostate cancer, melanoma, epidermoid carcinoma, sarcomas, bladder carcinomas, pancreatic cancers, kidney cancers, leukemias/lymphomas,

ovarian/uterine/cervical, gastric, colon, lung and CNS cancer cell lines. In addition, there are two independent samples of cerebellum. These cells are all cultured under standard recommended conditions and RNA extracted using the standard procedures. The cell lines in panel 3D, 3.1, 3.2, 1, 1.1., 1.2, 1.3D, 1.4, 1.5, and 1.6 are of the most common cell lines  
5 used in the scientific literature.

#### **AI.05 chondrosarcoma**

The AI.05 chondrosarcoma plates are comprised of SW1353 cells that had been subjected to serum starvation and treatment with cytokines that are known to induce MMP (1, 3 and 13) synthesis (eg. IL1beta). These treatments include: IL-1beta (10 ng/ml), IL-1beta +  
10 TNF-alpha (50 ng/ml), IL-1beta + Oncostatin (50 ng/ml) and PMA (100 ng/ml). The SW1353 cells were obtained from the ATCC (American Type Culture Collection) and were all cultured under standard recommended conditions. The SW1353 cells were plated at  $3 \times 10^5$  cells/ml (in DMEM medium-10 % FBS) in 6-well plates. The treatment was done in triplicate, for 6 and 18 h. The supernatants were collected for analysis of MMP 1, 3 and 13  
15 production and for RNA extraction. RNA was prepared from these samples using the standard procedures.

#### **Panels 4D, 4R, and 4.1D**

Panel 4 includes samples on a 96 well plate (2 control wells, 94 test samples) composed of RNA (Panel 4R) or cDNA (Panels 4D/4.1D) isolated from various human cell  
20 lines or tissues related to inflammatory conditions. Total RNA from control normal tissues such as colon and lung (Stratagene, La Jolla, CA) and thymus and kidney (Clontech) was employed. Total RNA from liver tissue from cirrhosis patients and kidney from lupus patients was obtained from BioChain (Biochain Institute, Inc., Hayward, CA). Intestinal tissue for RNA preparation from patients diagnosed as having Crohn's disease and  
25 ulcerative colitis was obtained from the National Disease Research Interchange (NDRI) (Philadelphia, PA).

Astrocytes, lung fibroblasts, dermal fibroblasts, coronary artery smooth muscle cells, small airway epithelium, bronchial epithelium, microvascular dermal endothelial cells, microvascular lung endothelial cells, human pulmonary aortic endothelial cells,  
30 human umbilical vein endothelial cells were all purchased from Clonetics (Walkersville, MD) and grown in the media supplied for these cell types by Clonetics. These primary cell types were activated with various cytokines or combinations of cytokines for 6 and/or

12-14 hours, as indicated. The following cytokines were used; IL-1 beta at approximately 1-5ng/ml, TNF alpha at approximately 5-10ng/ml, IFN gamma at approximately 20-50ng/ml, IL-4 at approximately 5-10ng/ml, IL-9 at approximately 5-10ng/ml, IL-13 at approximately 5-10ng/ml. Endothelial cells were sometimes starved for various times by culture in the basal media from Clonetics with 0.1% serum.

Mononuclear cells were prepared from blood of employees at CuraGen Corporation, using Ficoll. LAK cells were prepared from these cells by culture in DMEM 5% FCS (Hyclone), 100μM non essential amino acids (Gibco/Life Technologies, Rockville, MD), 1mM sodium pyruvate (Gibco), mercaptoethanol  $5.5 \times 10^{-5}$ M (Gibco), and 10mM Hepes (Gibco) and Interleukin 2 for 4-6 days. Cells were then either activated with 10-20ng/ml PMA and 1-2μg/ml ionomycin, IL-12 at 5-10ng/ml, IFN gamma at 20-50ng/ml and IL-18 at 5-10ng/ml for 6 hours. In some cases, mononuclear cells were cultured for 4-5 days in DMEM 5% FCS (Hyclone), 100μM non essential amino acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol  $5.5 \times 10^{-5}$ M (Gibco), and 10mM Hepes (Gibco) with PHA (phytohemagglutinin) or PWM (pokeweed mitogen) at approximately 5μg/ml. Samples were taken at 24, 48 and 72 hours for RNA preparation. MLR (mixed lymphocyte reaction) samples were obtained by taking blood from two donors, isolating the mononuclear cells using Ficoll and mixing the isolated mononuclear cells 1:1 at a final concentration of approximately  $2 \times 10^6$  cells/ml in DMEM 5% FCS (Hyclone), 100μM non essential amino acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol ( $5.5 \times 10^{-5}$ M) (Gibco), and 10mM Hepes (Gibco). The MLR was cultured and samples taken at various time points ranging from 1- 7 days for RNA preparation.

Monocytes were isolated from mononuclear cells using CD14 Miltenyi Beads, +ve VS selection columns and a Vario Magnet according to the manufacturer's instructions. Monocytes were differentiated into dendritic cells by culture in DMEM 5% fetal calf serum (FCS) (Hyclone, Logan, UT), 100μM non essential amino acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol  $5.5 \times 10^{-5}$ M (Gibco), and 10mM Hepes (Gibco), 50ng/ml GMCSF and 5ng/ml IL-4 for 5-7 days. Macrophages were prepared by culture of monocytes for 5-7 days in DMEM 5% FCS (Hyclone), 100μM non essential amino acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol  $5.5 \times 10^{-5}$ M (Gibco), 10mM Hepes (Gibco) and 10% AB Human Serum or MCSF at approximately 50ng/ml. Monocytes, macrophages and dendritic cells were stimulated for 6 and 12-14 hours with

lipopolysaccharide (LPS) at 100ng/ml. Dendritic cells were also stimulated with anti-CD40 monoclonal antibody (Pharmingen) at 10µg/ml for 6 and 12-14 hours.

CD4 lymphocytes, CD8 lymphocytes and NK cells were also isolated from mononuclear cells using CD4, CD8 and CD56 Miltenyi beads, positive VS selection columns and a Vario Magnet according to the manufacturer's instructions. CD45RA and CD45RO CD4 lymphocytes were isolated by depleting mononuclear cells of CD8, CD56, CD14 and CD19 cells using CD8, CD56, CD14 and CD19 Miltenyi beads and positive selection. CD45RO beads were then used to isolate the CD45RO CD4 lymphocytes with the remaining cells being CD45RA CD4 lymphocytes. CD45RA CD4, CD45RO CD4 and CD8 lymphocytes were placed in DMEM 5% FCS (Hyclone), 100µM non essential amino acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol  $5.5 \times 10^{-5}$ M (Gibco), and 10mM Hepes (Gibco) and plated at  $10^6$  cells/ml onto Falcon 6 well tissue culture plates that had been coated overnight with 0.5µg/ml anti-CD28 (Pharmingen) and 3µg/ml anti-CD3 (OKT3, ATCC) in PBS. After 6 and 24 hours, the cells were harvested for RNA preparation. To prepare chronically activated CD8 lymphocytes, we activated the isolated CD8 lymphocytes for 4 days on anti-CD28 and anti-CD3 coated plates and then harvested the cells and expanded them in DMEM 5% FCS (Hyclone), 100µM non essential amino acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol  $5.5 \times 10^{-5}$ M (Gibco), and 10mM Hepes (Gibco) and IL-2. The expanded CD8 cells were then activated again with plate bound anti-CD3 and anti-CD28 for 4 days and expanded as before. RNA was isolated 6 and 24 hours after the second activation and after 4 days of the second expansion culture. The isolated NK cells were cultured in DMEM 5% FCS (Hyclone), 100µM non essential amino acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol  $5.5 \times 10^{-5}$ M (Gibco), and 10mM Hepes (Gibco) and IL-2 for 4-6 days before RNA was prepared.

To obtain B cells, tonsils were procured from NDRI. The tonsil was cut up with sterile dissecting scissors and then passed through a sieve. Tonsil cells were then spun down and resuspended at  $10^6$  cells/ml in DMEM 5% FCS (Hyclone), 100µM non essential amino acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol  $5.5 \times 10^{-5}$ M (Gibco), and 10mM Hepes (Gibco). To activate the cells, we used PWM at 5µg/ml or anti-CD40 (Pharmingen) at approximately 10µg/ml and IL-4 at 5-10ng/ml. Cells were harvested for RNA preparation at 24, 48 and 72 hours.

To prepare the primary and secondary Th1/Th2 and Tr1 cells, six-well Falcon plates were coated overnight with 10µg/ml anti-CD28 (Pharmingen) and 2µg/ml OKT3 (ATCC),

and then washed twice with PBS. Umbilical cord blood CD4 lymphocytes (Poietic Systems, German Town, MD) were cultured at  $10^5$ - $10^6$  cells/ml in DMEM 5% FCS (Hyclone), 100 $\mu$ M non essential amino acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol  $5.5 \times 10^{-5}$ M (Gibco), 10mM Hepes (Gibco) and IL-2 (4ng/ml). IL-12 (5ng/ml) and anti-IL4 (1 $\mu$ g/ml) were used to direct to Th1, while IL-4 (5ng/ml) and anti-IFN gamma (1 $\mu$ g/ml) were used to direct to Th2 and IL-10 at 5ng/ml was used to direct to Tr1. After 4-5 days, the activated Th1, Th2 and Tr1 lymphocytes were washed once in DMEM and expanded for 4-7 days in DMEM 5% FCS (Hyclone), 100 $\mu$ M non essential amino acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol  $5.5 \times 10^{-5}$ M (Gibco), 10mM Hepes (Gibco) and IL-2 (1ng/ml). Following this, the activated Th1, Th2 and Tr1 lymphocytes were re-stimulated for 5 days with anti-CD28/OKT3 and cytokines as described above, but with the addition of anti-CD95L (1 $\mu$ g/ml) to prevent apoptosis. After 4-5 days, the Th1, Th2 and Tr1 lymphocytes were washed and then expanded again with IL-2 for 4-7 days. Activated Th1 and Th2 lymphocytes were maintained in this way for a maximum of three cycles. RNA was prepared from primary and secondary Th1, Th2 and Tr1 after 6 and 24 hours following the second and third activations with plate bound anti-CD3 and anti-CD28 mAbs and 4 days into the second and third expansion cultures in Interleukin 2.

The following leukocyte cells lines were obtained from the ATCC: Ramos, EOL-1, KU-812. EOL cells were further differentiated by culture in 0.1mM dbcAMP at  $5 \times 10^5$  cells/ml for 8 days, changing the media every 3 days and adjusting the cell concentration to  $5 \times 10^5$  cells/ml. For the culture of these cells, we used DMEM or RPMI (as recommended by the ATCC), with the addition of 5% FCS (Hyclone), 100 $\mu$ M non essential amino acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol  $5.5 \times 10^{-5}$ M (Gibco), 10mM Hepes (Gibco). RNA was either prepared from resting cells or cells activated with PMA at 10ng/ml and ionomycin at 1 $\mu$ g/ml for 6 and 14 hours. Keratinocyte line CCD106 and an airway epithelial tumor line NCI-H292 were also obtained from the ATCC. Both were cultured in DMEM 5% FCS (Hyclone), 100 $\mu$ M non essential amino acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol  $5.5 \times 10^{-5}$ M (Gibco), and 10mM Hepes (Gibco). CCD1106 cells were activated for 6 and 14 hours with approximately 5 ng/ml TNF alpha and 1ng/ml IL-1 beta, while NCI-H292 cells were activated for 6 and 14 hours with the following cytokines: 5ng/ml IL-4, 5ng/ml IL-9, 5ng/ml IL-13 and 25ng/ml IFN gamma.

For these cell lines and blood cells, RNA was prepared by lysing approximately  $10^7$  cells/ml using Trizol (Gibco BRL). Briefly, 1/10 volume of bromochloropropane (Molecular Research Corporation) was added to the RNA sample, vortexed and after 10 minutes at room temperature, the tubes were spun at 14,000 rpm in a Sorvall SS34 rotor.

5. The aqueous phase was removed and placed in a 15ml Falcon Tube. An equal volume of isopropanol was added and left at  $-20^{\circ}\text{C}$  overnight. The precipitated RNA was spun down at 9,000 rpm for 15 min in a Sorvall SS34 rotor and washed in 70% ethanol. The pellet was redissolved in 300 $\mu\text{l}$  of RNase-free water and 35 $\mu\text{l}$  buffer (Promega) 5 $\mu\text{l}$  DTT, 7 $\mu\text{l}$  RNasin and 8 $\mu\text{l}$  DNase were added. The tube was incubated at  $37^{\circ}\text{C}$  for 30 minutes to
- 10 remove contaminating genomic DNA, extracted once with phenol chloroform and re-precipitated with 1/10 volume of 3M sodium acetate and 2 volumes of 100% ethanol. The RNA was spun down and placed in RNase free water. RNA was stored at  $-80^{\circ}\text{C}$ .

#### AI\_comprehensive panel\_v1.0

- 15 The plates for AI\_comprehensive panel\_v1.0 include two control wells and 89 test samples comprised of cDNA isolated from surgical and postmortem human tissues obtained from the Backus Hospital and Clinomics (Frederick, MD). Total RNA was extracted from tissue samples from the Backus Hospital in the Facility at CuraGen. Total RNA from other tissues was obtained from Clinomics.

- 20 Joint tissues including synovial fluid, synovium, bone and cartilage were obtained from patients undergoing total knee or hip replacement surgery at the Backus Hospital. Tissue samples were immediately snap frozen in liquid nitrogen to ensure that isolated RNA was of optimal quality and not degraded. Additional samples of osteoarthritis and rheumatoid arthritis joint tissues were obtained from Clinomics. Normal control tissues
- 25 were supplied by Clinomics and were obtained during autopsy of trauma victims.

Surgical specimens of psoriatic tissues and adjacent matched tissues were provided as total RNA by Clinomics. Two male and two female patients were selected between the ages of 25 and 47. None of the patients were taking prescription drugs at the time samples were isolated.

- 30 Surgical specimens of diseased colon from patients with ulcerative colitis and Crohns disease and adjacent matched tissues were obtained from Clinomics. Bowel tissue from three female and three male Crohn's patients between the ages of 41-69 were used. Two patients were not on prescription medication while the others were taking

dexamethasone, phenobarbital, or tylenol. Ulcerative colitis tissue was from three male and four female patients. Four of the patients were taking lebid and two were on phenobarbital.

5 Total RNA from post mortem lung tissue from trauma victims with no disease or with emphysema, asthma or COPD was purchased from Clinomics. Emphysema patients ranged in age from 40-70 and all were smokers, this age range was chosen to focus on patients with cigarette-linked emphysema and to avoid those patients with alpha-1anti-trypsin deficiencies. Asthma patients ranged in age from 36-75, and excluded smokers to prevent those patients that could also have COPD. COPD patients ranged in age 10 from 35-80 and included both smokers and non-smokers. Most patients were taking corticosteroids, and bronchodilators.

In the labels employed to identify tissues in the AI\_comprehensive panel\_v1.0 panel, the following abbreviations are used:

15 AI = Autoimmunity  
Syn = Synovial  
Normal = No apparent disease  
Rep22 /Rep20 = individual patients  
RA = Rheumatoid arthritis  
Backus = From Backus Hospital  
20 OA = Osteoarthritis  
(SS) (BA) (MF) = Individual patients  
Adj = Adjacent tissue  
Match control = adjacent tissues  
-M = Male  
25 -F = Female  
COPD = Chronic obstructive pulmonary disease

#### **Panels 5D and 5I**

30 The plates for Panel 5D and 5I include two control wells and a variety of cDNAs isolated from human tissues and cell lines with an emphasis on metabolic diseases. Metabolic tissues were obtained from patients enrolled in the Gestational Diabetes study. Cells were obtained during different stages in the differentiation of adipocytes from human mesenchymal stem cells. Human pancreatic islets were also obtained.

In the Gestational Diabetes study subjects are young (18 - 40 years), otherwise healthy women with and without gestational diabetes undergoing routine (elective) Caesarean section. After delivery of the infant, when the surgical incisions were being repaired/closed, the obstetrician removed a small sample (<1 cc) of the exposed metabolic tissues during the closure of each surgical level. The biopsy material was rinsed in sterile saline, blotted and fast frozen within 5 minutes from the time of removal. The tissue was then flash frozen in liquid nitrogen and stored, individually, in sterile screw-top tubes and kept on dry ice for shipment to or to be picked up by CuraGen. The metabolic tissues of interest include uterine wall (smooth muscle), visceral adipose, skeletal muscle (rectus) and subcutaneous adipose. Patient descriptions are as follows:

5	Patient 2	Diabetic Hispanic, overweight, not on insulin
	Patient 7-9	Nondiabetic Caucasian and obese (BMI>30)
	Patient 10	Diabetic Hispanic, overweight, on insulin
	Patient 11	Nondiabetic African American and overweight
10	Patient 12	Diabetic Hispanic on insulin

Adipocyte differentiation was induced in donor progenitor cells obtained from Osirus (a division of Clonetics/BioWhittaker) in triplicate, except for Donor 3U which had only two replicates. Scientists at Clonetics isolated, grew and differentiated human mesenchymal stem cells (HuMSCs) for CuraGen based on the published protocol found in Mark F. Pittenger, et al., Multilineage Potential of Adult Human Mesenchymal Stem Cells Science Apr 2 1999: 143-147. Clonetics provided Trizol lysates or frozen pellets suitable for mRNA isolation and ds cDNA production. A general description of each donor is as follows:

25	Donor 2 and 3 U: Mesenchymal Stem cells, Undifferentiated Adipose
	Donor 2 and 3 AM: Adipose, AdiposeMidway Differentiated
	Donor 2 and 3 AD: Adipose, Adipose Differentiated

Human cell lines were generally obtained from ATCC (American Type Culture Collection), NCI or the German tumor cell bank and fall into the following tissue groups: kidney proximal convoluted tubule, uterine smooth muscle cells, small intestine, liver HepG2 cancer cells, heart primary stromal cells, and adrenal cortical adenoma cells. These cells are all cultured under standard recommended conditions and RNA extracted using the standard procedures. All samples were processed at CuraGen to produce single stranded cDNA.

Panel 5I contains all samples previously described with the addition of pancreatic islets from a 58 year old female patient obtained from the Diabetes Research Institute at the University of Miami School of Medicine. Islet tissue was processed to total RNA at an outside source and delivered to CuraGen for addition to panel 5I.

5 In the labels employed to identify tissues in the 5D and 5I panels, the following abbreviations are used:

GO Adipose = Greater Omentum Adipose

SK = Skeletal Muscle

UT = Uterus

10 PL = Placenta

AD = Adipose Differentiated

AM = Adipose Midway Differentiated

U = Undifferentiated Stem Cells

#### 15 **Panel CNSD.01**

The plates for Panel CNSD.01 include two control wells and 94 test samples comprised of cDNA isolated from postmortem human brain tissue obtained from the Harvard Brain Tissue Resource Center. Brains are removed from calvaria of donors between 4 and 24 hours after death, sectioned by neuroanatomists, and frozen at -80°C in liquid nitrogen vapor. All brains are sectioned and examined by neuropathologists to confirm diagnoses with clear associated neuropathology.

Disease diagnoses are taken from patient records. The panel contains two brains from each of the following diagnoses: Alzheimer's disease, Parkinson's disease, Huntington's disease, Progressive Supranuclear Palsy, Depression, and "Normal controls".

25 Within each of these brains, the following regions are represented: cingulate gyrus, temporal pole, globus pallidus, substantia nigra, Brodman Area 4 (primary motor strip), Brodman Area 7 (parietal cortex), Brodman Area 9 (prefrontal cortex), and Brodman area 17 (occipital cortex). Not all brain regions are represented in all cases; *e.g.*, Huntington's disease is characterized in part by neurodegeneration in the globus pallidus, thus this

30 region is impossible to obtain from confirmed Huntington's cases. Likewise Parkinson's disease is characterized by degeneration of the substantia nigra making this region more difficult to obtain. Normal control brains were examined for neuropathology and found to be free of any pathology consistent with neurodegeneration.

In the labels employed to identify tissues in the CNS panel, the following abbreviations are used:

PSP = Progressive supranuclear palsy

Sub Nigra = Substantia nigra

5 Glob. Palladus= Globus palladus

Temp Pole = Temporal pole

Cing Gyr = Cingulate gyrus

BA 4 = Brodman Area 4

**Panel CNS\_Neurodegeneration\_V1.0**

10 The plates for Panel CNS\_Neurodegeneration\_V1.0 include two control wells and  
47 test samples comprised of cDNA isolated from postmortem human brain tissue obtained  
from the Harvard Brain Tissue Resource Center (McLean Hospital) and the Human Brain  
and Spinal Fluid Resource Center (VA Greater Los Angeles Healthcare System). Brains are  
removed from calvaria of donors between 4 and 24 hours after death, sectioned by  
15 neuroanatomists, and frozen at -80°C in liquid nitrogen vapor. All brains are sectioned and  
examined by neuropathologists to confirm diagnoses with clear associated neuropathology.

Disease diagnoses are taken from patient records. The panel contains six brains  
from Alzheimer's disease (AD) patients, and eight brains from "Normal controls" who  
showed no evidence of dementia prior to death. The eight normal control brains are divided  
20 into two categories: Controls with no dementia and no Alzheimer's like pathology  
(Controls) and controls with no dementia but evidence of severe Alzheimer's like  
pathology, (specifically senile plaque load rated as level 3 on a scale of 0-3; 0 = no  
evidence of plaques, 3 = severe AD senile plaque load). Within each of these brains, the  
following regions are represented: hippocampus, temporal cortex (Brodman Area 21),  
25 parietal cortex (Brodman area 7), and occipital cortex (Brodman area 17). These regions  
were chosen to encompass all levels of neurodegeneration in AD. The hippocampus is a  
region of early and severe neuronal loss in AD; the temporal cortex is known to show  
neurodegeneration in AD after the hippocampus; the parietal cortex shows moderate  
neuronal death in the late stages of the disease; the occipital cortex is spared in AD and  
30 therefore acts as a "control" region within AD patients. Not all brain regions are  
represented in all cases.

In the labels employed to identify tissues in the CNS\_Neurodegeneration\_V1.0  
panel, the following abbreviations are used:

AD = Alzheimer's disease brain; patient was demented and showed AD-like pathology upon autopsy

Control = Control brains; patient not demented, showing no neuropathology

Control (Path) = Control brains; patient not demented but showing severe

5 AD-like pathology

SupTemporal Ctx = Superior Temporal Cortex

Inf Temporal Ctx = Inferior Temporal Cortex

#### A. CG101683-01: COT.

- 10 Expression of gene CG101683-01 was assessed using the primer-probe sets Ag3116, Ag3551 and Ag4828, described in Tables AA, AB and AC. Results of the RTQ-PCR runs are shown in Tables AD, AE, AF, AG, AH, AI and AJ.

Table AA. Probe Name Ag3116

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-catgttctcaagggacttgatt-3'	22	1072	453
Probe	TET-5'-cactcaaagaaagtgatccatcatga-3'-TAMRA	26	1099	454
Reverse	5'-ttttgtggacatgaaaacaatg-3'	22	1140	455

Table AB. Probe Name Ag3551

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-catgttctcaagggacttgatt-3'	22	1072	456
Probe	TET-5'-cactcaaagaaagtgatccatcatga-3'-TAMRA	26	1099	457
Reverse	5'-ttttgtggacatgaaaacaatg-3'	22	1140	458

- 15 Table AC. Probe Name Ag4828

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-gaggaatctgagatgctcaaga-3'	22	1663	459
Probe	TET-5'-caacgctctctctacatcgacctcgg-3'-TAMRA	26	1687	460
Reverse	5'-tccccgaacaagattgaagt-3'	20	1727	461

Table AD. CNS\_neurodegeneration\_v1.0

Tissue Name	Rel. Exp.(%) Ag3551, Run 209990366	Tissue Name	Rel. Exp.(%) Ag3551, Run 209990366
AD 1 Hippo	20.0	Control (Path) 3 Temporal Ctx	14.6
AD 2 Hippo	44.1	Control (Path) 4 Temporal Ctx	18.8
AD 3 Hippo	7.1	AD 1 Occipital Ctx	13.5
AD 4 Hippo	5.6	AD 2 Occipital Ctx (Missing)	0.0
AD 5 hippo	100.0	AD 3 Occipital Ctx	4.0
AD 6 Hippo	57.0	AD 4 Occipital Ctx	15.8
Control 2 Hippo	24.7	AD 5 Occipital Ctx	34.6
Control 4 Hippo	51.4	AD 6 Occipital Ctx	46.0
Control (Path) 3 Hippo	48.6	Control 1 Occipital Ctx	21.0
AD 1 Temporal Ctx	21.3	Control 2 Occipital Ctx	41.5
AD 2 Temporal Ctx	39.5	Control 3 Occipital Ctx	16.3
AD 3 Temporal Ctx	6.1	Control 4 Occipital Ctx	13.0
AD 4 Temporal Ctx	16.8	Control (Path) 1 Occipital Ctx	95.3
AD 5 Inf Temporal Ctx	100.0	Control (Path) 2 Occipital Ctx	10.2
AD 5 SupTemporal Ctx	91.4	Control (Path) 3 Occipital Ctx	21.5
AD 6 Inf Temporal Ctx	58.2	Control (Path) 4 Occipital Ctx	24.0
AD 6 Sup Temporal Ctx	65.5	Control 1 Parietal Ctx	17.2
Control 1 Temporal Ctx	20.3	Control 2 Parietal Ctx	57.4
Control 2 Temporal Ctx	21.2	Control 3 Parietal Ctx	16.5
Control 3 Temporal Ctx	10.8	Control (Path) 1 Parietal Ctx	28.3
Control 4 Temporal	6.9	Control (Path) 2	15.8

Ctx		Parietal Ctx	
Control (Path) 1. Temporal Ctx	42.0	Control (Path) 3 Parietal Ctx	19.6
Control (Path) 2 Temporal Ctx	26.4	Control (Path) 4 Parietal Ctx	61.1

Table AE. General\_screening\_panel\_v1.4

Tissue Name	Rel. Exp.(%) Ag3116, Run 219923407	Rel. Exp.(%) Ag3551, Run 218328114	Rel. Exp.(%) Ag4828, Run 217081802	Tissue Name	Rel. Exp.(%) Ag3116, Run 219923407	Rel. Exp.(%) Ag3551, Run 218328114	Rel. Exp.(%) Ag4828, Run 217081802
Adipose	100.0	58.2	53.6	Renal ca. TK-10	6.4	8.2	10.6
Melanoma* Hs688(A).T	18.8	9.0	15.5	Bladder	32.5	24.1	31.9
Melanoma* Hs688(B).T	21.3	10.7	17.4	Gastric ca. (liver met.) NCI-N87	26.8	23.5	36.3
Melanoma* M14	1.0	0.9	3.5	Gastric ca. KATO III	8.7	8.0	12.2
Melanoma* LOXIMVI	2.9	1.5	3.2	Colon ca. SW-948	2.6	2.6	5.4
Melanoma* SK-MEL-5	0.8	0.8	0.9	Colon ca. SW480	13.5	12.3	25.0
Squamous cell carcinoma SCC-4	1.0	2.2	7.0	Colon ca.* (SW480 met) SW620	1.6	1.4	2.5
Testis Pool	3.5	3.3	4.7	Colon ca. HT29	7.2	5.7	14.3
Prostate ca.* (bone met) PC-3	6.4	1.8	6.3	Colon ca. HCT-116	2.1	1.7	2.1
Prostate Pool	2.1	2.0	3.9	Colon ca. CaCo-2	13.5	15.7	15.9
Placenta	30.8	25.9	39.0	Colon cancer tissue	34.9	42.3	39.8
Uterus Pool	7.7	4.7	9.0	Colon ca. SW1116	0.1	0.3	3.4
Ovarian ca. OVCAR-3	4.4	6.1	15.7	Colon ca. Colo-205	2.7	2.6	8.8
Ovarian ca. SK-OV-3	9.7	18.2	46.3	Colon ca. SW-48	3.3	4.7	5.4
Ovarian ca.	3.7	5.4	7.1	Colon Pool	16.6	9.8	16.2

OVCAR-4							
Ovarian ca. OVCAR-5	19.2	19.9	30.6	Small Intestine Pool	7.3	5.5	9.3
Ovarian ca. IGROV-1	7.0	9.1	14.1	Stomach Pool	6.6	8.0	17.3
Ovarian ca. OVCAR-8	1.8	1.9	2.7	Bone Marrow Pool	5.2	3.3	7.0
Ovary	2.7	2.5	4.5	Fetal Heart	4.5	4.6	2.9
Breast ca. MCF-7	64.6	81.8	100.0	Heart Pool	9.2	6.8	7.9
Breast ca. MDA-MB- 231	3.1	2.1	9.2	Lymph Node Pool	10.4	9.9	15.2
Breast ca. BT 549	24.5	36.3	73.2	Fetal Skeletal Muscle	2.4	2.9	1.7
Breast ca. T47D	37.4	60.3	66.0 <sup>c</sup>	Skeletal Muscle Pool	7.7	8.5	9.8
Breast ca. MDA-N	0.3	0.5	0.9	Spleen Pool	16.0	22.8	45.7
Breast Pool	33.2	9.8	24.1	Thymus Pool	7.5	6.9	15.9
Trachea	14.5	15.5	18.0	CNS cancer (glio/astro) U87-MG	2.1	2.4	7.6
Lung	4.2	3.4	6.7	CNS cancer (glio/astro) U- 118-MG	5.4	2.7	7.9
Fetal Lung	83.5	100.0	68.3	CNS cancer (neuro;met) SK-N-AS	0.7	1.2	2.6
Lung ca. NCI-N417	0.0	0.0	0.2	CNS cancer (astro) SF-539	1.4	1.8	2.3
Lung ca. LX-1	8.0	6.0	11.8	CNS cancer (astro) SNB-75	4.7	5.9	14.1
Lung ca. NCI-H146	0.0	0.0	0.0	CNS cancer (glio) SNB-19	6.2	10.7	11.1
Lung ca. SHP-77	0.0	0.0	0.1	CNS cancer (glio) SF-295	16.0	18.8	31.9
Lung ca. A549	35.4	0.0	36.6	Brain (Amygdala) Pool	1.6	0.7	2.7
Lung ca. NCI-H526	0.0	0.0	0.0	Brain (cerebellum)	1.1	0.3	1.4
Lung ca. NCI-H23	10.9	13.0	13.4	Brain (fetal)	6.0	4.1	4.9
Lung ca.	7.4	5.8	17.6	Brain	3.6	1.5	3.7

NCI-H460				(Hippocampus) Pool			
Lung ca. HOP-62	11.4	4.3	13.2	Cerebral Cortex Pool	2.1	2.0	3.5
Lung ca. NCI-H522	1.6	1.5	2.1	Brain (Substantia nigra) Pool	2.4	2.0	2.7
Liver	0.6	0.2	1.0	Brain (Thalamus) Pool	2.6	2.2	4.5
Fetal Liver	5.0	4.0	2.8	Brain (whole)	2.7	2.5	4.5
Liver ca. HepG2	4.5	5.4	8.1	Spinal Cord Pool	2.1	3.2	3.8
Kidney Pool	26.6	21.0	31.4	Adrenal Gland	11.7	3.8	9.5
Fetal Kidney	9.0	10.7	7.7	Pituitary gland Pool	0.7	0.7	1.4
Renal ca. 786-0	6.0	7.9	10.9	Salivary Gland	1.9	1.5	2.5
Renal ca. A498	1.2	2.3	5.2	Thyroid (female)	3.3	3.6	7.7
Renal ca. ACHN	1.9	0.8	2.5	Pancreatic ca. CAPAN2	14.9	21.9	34.4
Renal ca. UO-31	11.1	10.7	14.9	Pancreas Pool	15.0	17.8	19.6

Table AF. Panel 1.3D

Tissue Name	Rel. Exp.(%) Ag3116, Run 167617379	Tissue Name	Rel. Exp.(%) Ag3116, Run 167617379
Liver adenocarcinoma	24.8	Kidney (fetal)	34.2
Pancreas	3.4	Renal ca. 786-0	3.7
Pancreatic ca. CAPAN 2	12.1	Renal ca. A498	3.3
Adrenal gland	2.6	Renal ca. RXF 393	17.1
Thyroid	1.3	Renal ca. ACHN	1.7
Salivary gland	0.0	Renal ca. UO-31	0.8
Pituitary gland	2.1	Renal ca. TK-10	4.4
Brain (fetal)	3.1	Liver	2.4
Brain (whole)	3.1	Liver (fetal)	4.5
Brain (amygdala)	1.0	Liver ca. (hepatoblast) HepG2	4.4
Brain (cerebellum)	1.0	Lung	25.0

Brain (hippocampus)	3.0	Lung (fetal)	29.7
Brain (substantia nigra)	3.7	Lung ca. (small cell) LX-1	5.5
Brain (thalamus)	1.2	Lung ca. (small cell) NCI-H69	0.0
Cerebral Cortex	2.5	Lung ca. (s.cell var.) SHP-77	0.0
Spinal cord	3.0	Lung ca. (large cell)NCI-H460	2.3
glio/astro U87-MG	1.5	Lung ca. (non-sm. cell) A549	14.3
glio/astro U-118-MG	2.8	Lung ca. (non-s.cell) NCI-H23	5.0
astrocytoma SW1783	2.0	Lung ca. (non-s.cell) HOP-62	5.7
neuro*; met SK-N-AS	1.5	Lung ca. (non-s.cl) NCI-H522	1.2
astrocytoma SF-539	2.4	Lung ca. (squam.) SW 900	24.1
astrocytoma SNB-75	14.5	Lung ca. (squam.) NCI-H596	0.0
glioma SNB-19	0.0	Mammary gland	7.7
glioma U251	0.7	Breast ca.* (pl.ef) MCF-7	57.8
glioma SF-295	6.9	Breast ca.* (pl.ef) MDA-MB-231	0.8
Heart (fetal)	5.8	Breast ca.* (pl.ef) T47D	3.5
Heart	3.2	Breast ca. BT-549	4.8
Skeletal muscle (fetal)	4.6	Breast ca. MDA-N	0.0
Skeletal muscle	2.1	Ovary	6.1
Bone marrow	4.0	Ovarian ca. OVCAR-3	3.0
Thymus	3.4	Ovarian ca. OVCAR-4	26.1
Spleen	10.6	Ovarian ca. OVCAR-5	44.8
Lymph node	10.3	Ovarian ca. OVCAR-8	1.4
Colorectal	6.4	Ovarian ca. IGROV- 1	6.4
Stomach	1.8	Ovarian ca.* (ascites) SK-OV-3	33.2
Small intestine	3.0	Uterus	4.4

Colon ca. SW480	6.0	Placenta	6.8
Colon ca.* SW620(SW480 met)	6.1	Prostate	0.0
Colon ca. HT29	6.6	Prostate ca.* (bone met)PC-3	2.1
Colon ca. HCT-116	0.0	Testis	0.0
Colon ca. CaCo-2	11.3	Melanoma Hs688(A).T	1.0
Colon ca. tissue(ODO3866)	13.1	Melanoma* (met) Hs688(B).T	3.5
Colon ca. HCC-2998	17.6	Melanoma UACC- 62	0.0
Gastric ca.* (liver met) NCI-N87	11.0	Melanoma M14	1.1
Bladder	10.2	Melanoma LOX IMVI	1.2
Trachea	3.9	Melanoma* (met) SK-MEL-5	0.0
Kidney	5.0	Adipose	100.0

Table AG. Panel 2D

Tissue Name	Rel. Exp.(%) Ag3116, Run 169556216	Tissue Name	Rel. Exp.(%) Ag3116, Run 169556216
Normal Colon	58.2	Kidney Margin 8120608	2.0
CC Well to Mod Diff (ODO3866)	22.7	Kidney Cancer 8120613	3.5
CC Margin (ODO3866)	14.4	Kidney Margin 8120614	2.9
CC Gr.2 rectosigmoid (ODO3868)	7.5	Kidney Cancer 9010320	42.0
CC Margin (ODO3868)	3.4	Kidney Margin 9010321	7.7
CC Mod Diff (ODO3920)	7.0	Normal Uterus	7.0
CC Margin (ODO3920)	6.9	Uterus Cancer 064011	18.8
CC Gr.2 ascend colon (ODO3921)	27.7	Normal Thyroid	5.8
CC Margin (ODO3921)	8.4	Thyroid Cancer 064010	6.9
CC from Partial Hepatectomy	34.9	Thyroid Cancer A302152	3.0

(ODO4309) Mets			
Liver Margin (ODO4309)	8.5	Thyroid Margin A302153	12.1
Colon mets to lung (OD04451-01)	12.2	Normal Breast	28.9
Lung Margin (OD04451-02)	21.8	Breast Cancer (OD04566)	6.3
Normal Prostate 6546-1	2.9	Breast Cancer (OD04590-01)	44.4
Prostate Cancer (OD04410)	7.4	Breast Cancer Mets (OD04590-03)	43.5
Prostate Margin (OD04410)	8.2	Breast Cancer Metastasis (OD04655-05)	6.9
Prostate Cancer (OD04720-01)	6.6	Breast Cancer 064006	12.0
Prostate Margin (OD04720-02)	21.8	Breast Cancer 1024	12.9
Normal Lung 061010	42.6	Breast Cancer 9100266	6.9
Lung Met to Muscle (ODO4286)	15.0	Breast Margin 9100265	6.9
Muscle Margin (ODO4286)	9.5	Breast Cancer A209073	7.2
Lung Malignant Cancer (OD03126)	17.4	Breast Margin A209073	4.3
Lung Margin (OD03126)	59.5	Normal Liver	2.3
Lung Cancer (OD04404)	53.6	Liver Cancer 064003	2.1
Lung Margin (OD04404)	45.1	Liver Cancer 1025	5.8
Lung Cancer (OD04565)	10.4	Liver Cancer 1026	4.2
Lung Margin (OD04565)	10.8	Liver Cancer 6004-T	6.1
Lung Cancer (OD04237-01)	39.8	Liver Tissue 6004-N	6.4
Lung Margin (OD04237-02)	65.5	Liver Cancer 6005-T	7.4
Ocular Mel Met to Liver (ODO4310)	1.6	Liver Tissue 6005-N	3.9
Liver Margin (ODO4310)	9.9	Normal Bladder	37.1
Melanoma Mets to Lung (OD04321)	2.0	Bladder Cancer 1023	6.5
Lung Margin (OD04321)	50.7	Bladder Cancer A302173	14.8
Normal Kidney	13.0	Bladder Cancer	27.9

		(OD04718-01)	
Kidney Ca, Nuclear grade 2 (OD04338)	16.4	Bladder Normal Adjacent (OD04718-03)	100.0
Kidney Margin (OD04338)	18.4	Normal Ovary	6.3
Kidney Ca Nuclear grade 1/2 (OD04339)	10.3	Ovarian Cancer 064008	31.9
Kidney Margin (OD04339)	6.5	Ovarian Cancer (OD04768-07)	21.9
Kidney Ca, Clear cell type (OD04340)	28.7	Ovary Margin (OD04768-08)	32.5
Kidney Margin (OD04340)	22.7	Normal Stomach	18.8
Kidney Ca, Nuclear grade 3 (OD04348)	4.5	Gastric Cancer 9060358	14.6
Kidney Margin (OD04348)	6.7	Stomach Margin 9060359	16.2
Kidney Cancer (OD04622-01)	12.2	Gastric Cancer 9060395	33.2
Kidney Margin (OD04622-03)	1.8	Stomach Margin 9060394	24.8
Kidney Cancer (OD04450-01)	4.0	Gastric Cancer 9060397	26.8
Kidney Margin (OD04450-03)	7.1	Stomach Margin 9060396	7.4
Kidney Cancer 8120607	3.3	Gastric Cancer 064005	27.4

Table AH. Panel 4D

Tissue Name	Rel. Exp.(%) Ag3116, Run 164526105	Rel. Exp.(%) Ag3551, Run 166453851	Tissue Name	Rel. Exp.(%) Ag3116, Run 164526105	Rel. Exp.(%) Ag3551, Run 166453851
Secondary Th1 act	15.6	38.4	HUVEC IL-1beta	0.8	8.2
Secondary Th2 act	23.0	56.3	HUVEC IFN gamma	1.4	1.2
Secondary Tr1 act	23.2	78.5	HUVEC TNF alpha + IFN gamma	3.0	3.1
Secondary Th1 rest	2.9	22.8	HUVEC TNF alpha + IL4	2.5	2.6
Secondary Th2 rest	2.5	4.5	HUVEC IL-11	0.5	0.5

Secondary Tr1 rest	2.0	7.0	Lung Microvascular EC none	0.0	0.1
Primary Th1 act	13.5	18.3	Lung Microvascular EC TNFalpha + IL- 1beta	4.2	2.8
Primary Th2 act	6.6	15.5	Microvascular Dermal EC none	0.1	0.1
Primary Tr1 act	17.7	33.2	Microvascular Dermal EC TNFalpha + IL- 1beta	5.7	7.3
Primary Th1 rest	9.2	32.1	Bronchial epithelium TNFalpha + IL1beta	2.4	1.5
Primary Th2 rest	1.2	2.9	Small airway epithelium none	0.6	1.1
Primary Tr1 rest	1.7	3.8	Small airway epithelium TNFalpha + IL- 1beta	5.5	5.0
CD45RA CD4 lymphocyte act	4.9	6.7	Coronary artery SMC rest	1.0	0.8
CD45RO CD4 lymphocyte act	11.1	44.8	Coronary artery SMC TNFalpha + IL-1beta	0.7	0.6
CD8 lymphocyte act	5.3	12.2	Astrocytes rest	0.5	1.0
Secondary CD8 lymphocyte rest	4.9	16.0	Astrocytes TNFalpha + IL- 1beta	14.9	61.1
Secondary CD8 lymphocyte act	7.6	25.5	KU-812 (Basophil) rest	0.2	0.2
CD4 lymphocyte none	0.8	1.1	KU-812 (Basophil) PMA/ionomycin	1.0	1.5
2ry Th1/Th2/Tr1_anti- CD95 CH11	3.0	11.0	CCD1106 (Keratinocytes) none	0.4	0.5
LAK cells rest	6.8	5.3	CCD1106 (Keratinocytes) TNFalpha + IL- 1beta	0.8	12.4
LAK cells IL-2	6.4	23.2	Liver cirrhosis	1.1	5.3

LAK cells IL-2+IL-12	22.4	73.7	Lupus kidney	1.1	4.8
LAK cells IL-2+IFN gamma	17.4	44.1	NCI-H292 none	8.4	9.7
LAK cells IL-2+IL-18	12.2	25.0	NCI-H292 IL-4	17.6	18.4
LAK cells PMA/ionomycin	12.3	20.7	NCI-H292 IL-9	6.5	5.3
NK Cells IL-2 rest	12.9	23.0	NCI-H292 IL-13	9.2	12.0
Two Way MLR 3 day	12.5	24.0	NCI-H292 IFN gamma	4.3	3.5
Two Way MLR 5 day	6.0	17.1	HPAEC none	0.5	0.5
Two Way MLR 7 day	3.0	6.3	HPAEC TNF alpha + IL-1 beta	8.2	11.0
PBMC rest	4.0	5.4	Lung fibroblast none	0.2	1.0
PBMC PWM	100.0	49.3	Lung fibroblast TNF alpha + IL-1 beta	1.7	9.8
PBMC PHA-L	11.8	5.6	Lung fibroblast IL-4	3.3	3.2
Ramos (B cell) none	0.8	2.0	Lung fibroblast IL-9	0.9	0.5
Ramos (B cell) ionomycin	16.7	6.5	Lung fibroblast IL-13	1.4	1.8
B lymphocytes PWM	53.2	25.3	Lung fibroblast IFN gamma	3.4	4.0
B lymphocytes CD40L and IL-4	61.1	81.8	Dermal fibroblast CCD1070 rest	1.9	1.1
EOL-1 dbcAMP	0.7	0.4	Dermal fibroblast CCD1070 TNF alpha	11.9	13.7
EOL-1 dbcAMP PMA/ionomycin	2.2	3.0	Dermal fibroblast CCD1070 IL-1 beta	6.1	6.3
Dendritic cells none	4.8	8.7	Dermal fibroblast IFN gamma	0.6	0.9
Dendritic cells LPS	12.3	25.2	Dermal fibroblast IL-4	4.2	6.7
Dendritic cells anti-CD40	3.2	6.8	IBD Colitis 2	1.1	4.1
Monocytes rest	5.0	7.3	IBD Crohn's	1.8	6.0
Monocytes LPS	43.8	100.0	Colon	2.6	15.7
Macrophages rest	8.2	11.7	Lung	8.2	7.5

Macrophages LPS	26.8	57.4	Thymus	2.3	3.5
HUVEC none	0.2	0.5	Kidney	4.2	3.8
HUVEC starved	0.6	1.5			

Table AI. Panel 5D

Tissue Name	Rel. Exp.(%) Ag3116, Run 170863008	Rel. Exp.(%) Ag4828, Run 219436967	Tissue Name	Rel. Exp.(%) Ag3116, Run 170863008	Rel. Exp.(%) Ag4828, Run 219436967
97457_Patient-02go_adipose	33.4	33.9	94709_Donor 2 AM - A_adipose	5.1	10.8
97476_Patient-07sk_skeletal muscle	31.2	33.4	94710_Donor 2 AM - B_adipose	3.2	9.3
97477_Patient-07ut_uterus	7.7	59.5	94711_Donor 2 AM - C_adipose	0.0	3.0
97478_Patient-07pl_placenta	62.0	39.8	94712_Donor 2 AD - A_adipose	12.9	13.7
97481_Patient-08sk_skeletal muscle	20.0	25.9	94713_Donor 2 AD - B_adipose	12.9	10.0
97482_Patient-08ut_uterus	33.4	19.8	94714_Donor 2 AD - C_adipose	8.8	6.7
97483_Patient-08pl_placenta	58.6	41.5	94742_Donor 3 U - A_Mesenchymal Stem Cells	1.6	4.7
97486_Patient-09sk_skeletal muscle	3.7	6.5	94743_Donor 3 U - B_Mesenchymal Stem Cells	4.8	2.8
97487_Patient-09ut_uterus	13.6	8.1	94730_Donor 3 AM - A_adipose	6.8	6.3
97488_Patient-09pl_placenta	41.2	38.4	94731_Donor 3 AM - B_adipose	5.3	2.4
97492_Patient-10ut_uterus	31.9	30.6	94732_Donor 3 AM - C_adipose	1.9	2.2
97493_Patient-10pl_placenta	74.7	72.7	94733_Donor 3 AD - A_adipose	2.5	10.2
97495_Patient-11go_adipose	67.4	100.0	94734_Donor 3 AD - B_adipose	2.9	5.5
97496_Patient-11sk_skeletal muscle	9.0	5.8	94735_Donor 3 AD - C_adipose	6.7	4.7
97497_Patient-11ut_uterus	35.4	20.6	77138_Liver_HepG2untreated	13.0	14.4

97498_Patient-11pl_placenta	52.1	50.0	73556_Heart_Cardiac stromal cells (primary)	9.1	1.9
97500_Patient-12go_adipose	100.0	82.4	81735_Small Intestine	20.0	17.2
97501_Patient-12sk_skeletal muscle	14.2	19.2	72409_Kidney_Proximal Convoluted Tubule	0.0	0.9
97502_Patient-12ut_uterus	51.8	23.7	82685_Small intestine_Duodenum	13.5	19.1
97503_Patient-12pl_placenta	39.5	57.0	90650_Adrenal_Adrenocortical adenoma	7.3	8.8
94721_Donor 2 U - A_Mesenchymal Stem Cells	2.1	1.6	72410_Kidney_HRCE	9.9	7.6
94722_Donor 2 U - B_Mesenchymal Stem Cells	0.0	3.0	72411_Kidney_HRE	5.9	13.5
94723_Donor 2 U - C_Mesenchymal Stem Cells	1.8	2.1	73139_Uterus_Uterine smooth muscle cells	2.5	2.0

Table AJ. general oncology screening panel\_v\_2.4

Tissue Name	Rel. Exp.(%) Ag3551, Run 259737946	Tissue Name	Rel. Exp.(%) Ag3551, Run 259737946
Colon cancer 1	26.6	Bladder NAT 2	0.0
Colon NAT 1	9.4	Bladder NAT 3	1.5
Colon cancer 2	32.3	Bladder NAT 4	5.8
Colon NAT 2	7.1	Prostate adenocarcinoma 1	29.9
Colon cancer 3	69.3	Prostate adenocarcinoma 2	1.5
Colon NAT 3	41.5	Prostate adenocarcinoma 3	2.9
Colon malignant cancer 4	96.6	Prostate adenocarcinoma 4	69.3
Colon NAT 4	5.6	Prostate NAT 5	1.3
Lung cancer 1	34.6	Prostate adenocarcinoma 6	2.1
Lung NAT 1	5.4	Prostate adenocarcinoma 7	5.5
Lung cancer 2	100.0	Prostate	1.5

		adenocarcinoma 8	
Lung NAT 2	15.0	Prostate adenocarcinoma 9	19.1
Squamous cell carcinoma 3	37.6	Prostate NAT 10	0.0
Lung NAT 3	2.8	Kidney cancer 1	38.2
Metastatic melanoma 1	43.8	Kidney NAT 1	13.9
Melanoma 2	5.0	Kidney cancer 2	66.9
Melanoma 3	2.4	Kidney NAT 2	19.3
Metastatic melanoma 4	69.3	Kidney cancer 3	27.2
Metastatic melanoma 5	93.3	Kidney NAT 3	12.1
Bladder cancer 1	2.2	Kidney cancer 4	20.4
Bladder NAT 1	0.0	Kidney NAT 4	6.3
Bladder cancer 2	5.0		

**CNS\_neurodegeneration\_v1.0 Summary:** Ag3551 This panel confirms the expression of this gene at low levels in the brains of an independent group of individuals. However, no differential expression of this gene was detected between Alzheimer's diseased postmortem brains and those of non-demented controls in this experiment. Please see Panel 1.4 for a discussion of the potential utility of this gene in treatment of central nervous system disorders.

**General\_screening\_panel\_v1.4 Summary:** Ag3116/Ag3551/Ag4828 Results of three experiments with two different probes and primer sets are in excellent agreement. Highest expression of this gene is detected in adipose, fetal lung, and breast cancer MCF-7 cell lines (CTs=27-30). Interestingly, this gene is expressed at much higher levels in fetal (CTs=27-30) when compared to adult lung (CT =31-35). This observation suggests that expression of this gene can be used to distinguish fetal from adult lung. In addition, the relative overexpression of this gene in fetal lung suggests that the protein product may enhance lung growth or development in the fetus and thus may also act in a regenerative capacity in the adult. Therefore, therapeutic modulation of the protein encoded by this gene could be useful in treatment of lung related diseases.

In addition significant expression of this gene is found in a number of cancer (pancreatic, CNS, colon, lung, breast, ovary, prostate, melanoma) cell lines. Therefore, therapeutic

modulation of the activity of this gene or its protein product, through the use of small molecule drugs, might be beneficial in the treatment of these cancers.

Among tissues with metabolic or endocrine function, this gene is expressed at high to moderate levels in pancreas, adipose, adrenal gland, thyroid, skeletal muscle, heart, fetal  
5 liver and the gastrointestinal tract. Therefore, therapeutic modulation of the activity of this gene may prove useful in the treatment of endocrine/metabolically related diseases, such as obesity and diabetes..

This gene encodes a protein that is homologous to mitogen-activated protein kinase kinase kinase 8 (MAP3K8)(COT proto-oncogene serine/threonine-protein kinase) (C-COT)  
10 (Cancer osaka thyroid oncogene). COT is able to enhance the TNF alpha production and to activate NF-kB. Both events are connected with insulin resistance and type II diabetes (1, 2, 3). Inhibition of COT kinase would prevent overproduction of TNF alpha and activation of NF-kB, thus improving insulin resistance and diabetes.

In addition, this gene is expressed at high levels in all regions of the central nervous system  
15 examined, including amygdala, hippocampus, substantia nigra, thalamus, cerebellum, cerebral cortex, and spinal cord. Recently, MKK6, a related protein, has been shown to associated with Alzheimer's disease (4). Therefore, based on the homology of this protein to MKK6 and the presence of this gene in the brain, we predict that this putative MAP3K8 may play a role in central nervous system disorders such as Alzheimer's disease,  
20 Parkinson's disease, epilepsy, multiple sclerosis, schizophrenia and depression.

Ag3551 Results from one experiment (run 213391203) are not included. The amp plot indicates that there were experimental difficulties with this run. (Data not shown).

#### References:

1. Ballester A, Velasco A, Tobena R, Alemany S. Cot kinase activates tumor necrosis  
25 factor-alpha gene expression in a cyclosporin A-resistant manner. J. Biol. Chem. 1998. 273, 14099-106. PMID: 9603908.
2. Bierhaus A, Schiekofe S, Schwaninger M, Andrassy M, Humpert PM, Chen J, Hong M, Luther T, Henle T, Kloting I, Morcos M, Hofmann M, Tritschler H, Weigle B, Kasper M, Smith M, Perry G, Schmidt AM, Stern DM, Haring HU, Schleicher E, Nawroth PP.

Diabetes-associated sustained activation of the transcription factor nuclear factor-kappaB. Diabetes, 2001 50, 2792-808. PMID: 11723063.

3. Belich MP, Salmeron A, Johnston LH, Ley SC. TPL-2 kinase regulates the proteolysis of the NF-kappaB-inhibitory protein NF-kappaB1 p105. Nature. 1999 397, 363-8. PMID: 9950430.

4. Zhu X, Rottkamp CA, Hartzler A, Sun Z, Takeda A, Boux H, Shimohama S, Perry G, Smith MA. (2001) Activation of MKK6, an upstream activator of p38, in Alzheimer's disease. J Neurochem 79(2):311-8

**Panel 1.3D Summary:** Ag3116 Highest expression of this gene is detected in adipose (32.7). Low to moderate expression of this gene is also seen in number of ovarian cancer cell lines, liver adenocarcinoma and breast cancer MCF-7 cell line. Therefore, therapeutic modulation of the activity of this gene or its protein product, through the use of small molecule drugs, might be beneficial in the treatment of these cancers.

In addition, low expression of this gene is also seen in fetal kidney and lung. Interestingly, this gene is expressed at much higher levels in fetal (CT=34.3) when compared to adult kidney (CT=37). This observation suggests that expression of this gene can be used to distinguish fetal from adult kidney. In addition, the relative overexpression of this gene in fetal lung suggests that the protein product may enhance lung growth or development in the fetus and thus may also act in a regenerative capacity in the adult. Therefore, therapeutic modulation of the protein encoded by this gene could be useful in treatment of lung related diseases.

**Panel 2D Summary:** Ag3116 Highest expression of this gene is detected in normal bladder (OD04718-03) sample (CT=31.4). Low to moderate expression of this gene is seen in large number of normal and cancer samples. Please see Panel 1.4 for a discussion of the potential utility of this gene.

**Panel 4D Summary:** Ag3116/Ag3551 Results from two experiments with same primer and probe set are in excellent agreement. Highest expression of this gene is detected in PWM treated PBMC and LPS treated monocytes (CTs=28-29). Interestingly, expression of this gene is stimulated in activated primary Th2 and Tr1, activated secondary Th1, Th2,

Tr1, PWM treated PBMC, LPS treated monocytes, TNFalpha + IL-1beta treated astrocytes and keratinocytes. Thus, expression of this gene can be used to distinguish between these activated or treated cells from the corresponding untreated or resting cells.

- In addition low expression of this gene is seen in a wide range of cell types of significance in the immune response in health and disease. These cells include members of the T-cell, B-cell, endothelial cell, macrophage/monocyte, and peripheral blood mononuclear cell family, as well as epithelial and fibroblast cell types from lung and skin, and normal tissues represented by colon, lung, thymus and kidney. Therefore, modulation of the gene product with a functional therapeutic may lead to the alteration of functions associated with these cell types and lead to improvement of the symptoms of patients suffering from autoimmune and inflammatory diseases such as asthma, allergies, inflammatory bowel disease, lupus erythematosus, psoriasis, rheumatoid arthritis, and osteoarthritis.

- Panel 5D Summary:** Ag3116/Ag4828 Results from two experiments with different primer and probe set are in excellent agreement. Highest expression of this gene is detected in adipose tissue (CTs=29-33). Low to moderate expression of this gene is seen in wide range of samples used in this panel including adipose, skeletal muscle, uterus, and placenta. This wide spread expression of this gene in tissues with metabolic or endocrine function, suggests that this gene plays a role in endocrine/metabolically related diseases, such as obesity and diabetes.

- This gene codes for mitogen-activated protein kinase kinase kinase 8 (MAP3K8). Recently, activation of MAP kinase, ERK, a related protein, by modified LDL in vascular smooth muscle cells has been implicated in the development of atherosclerosis in diabetes (Ref.1). Therefore, MAP3K8 may also play a role in the development of this disease and therapeutic modulation of the activity of this gene or its protein product, through the use of small molecule drugs, might be beneficial in the treatment of atherosclerosis and diabetes.

#### References.

1. Velarde V, Jenkins AJ, Christopher J, Lyons TJ, Jaffa AA. (2001) Activation of MAPK by modified low-density lipoproteins in vascular smooth muscle cells. *J Appl Physiol* 91(3):1412-20. PMID: 11509543.

- General oncology screening panel\_v\_2.4 Summary:** Ag3551 Highest expression of this gene is detected in lung cancer (CT=32.3). Moderate to low expression of this gene is detected in metastatic melanoma, prostate, lung and kidney cancers. Interestingly, expression of this gene is higher in cancer as compared to normal tissues. Therefore,
- 5 expression of this gene may be used as diagnostic marker to detect the presence of these cancers and therapeutic modulation of this gene through the use of antibodies or small molecule may be useful in the treatment of metastatic melanoma, prostate, lung and kidney cancers.

**B. CG101996-02: Phosphorylase kinase gamma full length.**

- 10 Expression of gene CG101996-02 was assessed using the primer-probe sets Ag3882 and Ag5945, described in Tables BA and BB. Results of the RTQ-PCR runs are shown in Tables BC, BD, BE, BF and BG.

Table BA. Probe Name Ag3882

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-ctgatgctgaggatgatcatg-3'	21	828	462
Probe	TET-5'-aactaccagtttggtcgccccaggt-3'-TAMRA	25	855	463
Reverse	5'-cttcacggtgtccgagtaatc-3'	21	885	464

Table BB. Probe Name Ag5945

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-attcttgtcaagctccttcaaga-3'	23	45	465
Probe	TET-5'-caagcacttaaccagccaccagaggt-3'-TAMRA	26	73	466
Reverse	5'-gtcatgctcagatcttcagtga-3'	22	103	467

- 15 Table BC. AI\_comprehensive panel\_v1.0

Tissue Name	Rel. Exp.(%) Ag5945, Run 248201924	Tissue Name	Rel. Exp.(%) Ag5945, Run 248201924
110967 COPD-F	0.8	112427 Match Control Psoriasis-F	4.7
110980 COPD-F	3.8	112418 Psoriasis-M	8.1
110968 COPD-M	1.0	112723 Match Control	0.0

		Psoriasis-M	
110977 COPD-M	6.4	112419 Psoriasis-M	1.4
110989 Emphysema-F	0.4	112424 Match Control Psoriasis-M	0.0
110992 Emphysema-F	1.9	112420 Psoriasis-M	3.4
110993 Emphysema-F	1.2	112425 Match Control Psoriasis-M	5.1
110994 Emphysema-F	0.0	104689 (MF) OA Bone-Backus	55.5
110995 Emphysema-F	2.7	104690 (MF) Adj "Normal" Bone-Backus	72.7
110996 Emphysema-F	0.0	104691 (MF) OA Synovium-Backus	41.5
110997 Asthma-M	0.0	104692 (BA) OA Cartilage-Backus	30.8
111001 Asthma-F	1.5	104694 (BA) OA Bone-Backus	20.3
111002 Asthma-F	1.1	104695 (BA) Adj "Normal" Bone-Backus	69.3
111003 Atopic Asthma-F	0.4	104696 (BA) OA Synovium-Backus	14.3
111004 Atopic Asthma-F	0.4	104700 (SS) OA Bone-Backus	24.1
111005 Atopic Asthma-F	0.0	104701 (SS) Adj "Normal" Bone-Backus	51.4
111006 Atopic Asthma-F	0.3	104702 (SS) OA Synovium-Backus	64.2
111417 Allergy-M	0.2	117093 OA Cartilage Rep7	0.2
112347 Allergy-M	0.3	112672 OA Bone5	5.9
112349 Normal Lung-F	0.6	112673 OA Synovium5	3.9
112357 Normal Lung-F	1.7	112674 OA Synovial Fluid cells5	0.2
112354 Normal Lung-M	2.5	117100 OA Cartilage Rep14	0.1
112374 Crohns-F	0.9	112756 OA Bone9	0.0
112389 Match Control Crohns-F	1.2	112757 OA Synovium9	100.0
112375 Crohns-F	2.8	112758 OA Synovial Fluid Cells9	0.7

112732 Match Control Crohns-F	1.9	117125 RA Cartilage Rep2	0.7
112725 Crohns-M	0.0	113492 Bone2 RA	3.2
112387 Match Control Crohns-M	0.4	113493 Synovium2 RA	1.8
112378 Crohns-M	0.1	113494 Syn Fluid Cells RA	1.5
112390 Match Control Crohns-M	3.2	113499 Cartilage4 RA	2.8
112726 Crohns-M	0.6	113500 Bone4 RA	1.1
112731 Match Control Crohns-M	1.2	113501 Synovium4 RA	0.9
112380 Ulcer Col-F	0.0	113502 Syn Fluid Cells4 RA	0.6
112734 Match Control Ulcer Col-F	1.9	113495 Cartilage3 RA	2.5
112384 Ulcer Col-F	0.9	113496 Bone3 RA	2.1
112737 Match Control Ulcer Col-F	0.4	113497 Synovium3 RA	1.6
112386 Ulcer Col-F	0.0	113498 Syn Fluid Cells3 RA	2.1
112738 Match Control Ulcer Col-F	2.6	117106 Normal Cartilage Rep20	0.0
112381 Ulcer Col-M	0.0	113663 Bone3 Normal	0.5
112735 Match Control Ulcer Col-M	1.4	113664 Synovium3 Normal	0.0
112382 Ulcer Col-M	0.8	113665 Syn Fluid Cells3 Normal	0.0
112394 Match Control Ulcer Col-M	0.3	117107 Normal Cartilage Rep22	0.8
112383 Ulcer Col-M	0.0	113667 Bone4 Normal	0.1
112736 Match Control Ulcer Col-M	0.4	113668 Synovium4 Normal	1.5
112423 Psoriasis-F	0.4	113669 Syn Fluid Cells4 Normal	0.8

Table BD. General\_screening\_panel\_v1.4

Tissue Name	Rel. Exp.(%) Ag3882, Run 217334262	Rel. Exp.(%) Ag3882, Run 222181244	Rel. Exp.(%) Ag3882, Run 222185729	Tissue Name	Rel. Exp.(%) Ag3882, Run 217334262	Rel. Exp.(%) Ag3882, Run 222181244	Rel. Exp.(%) Ag3882, Run 222185729
Adipose	2.1	3.9	2.5	Renal ca. TK- 10	2.8	2.4	3.8

Melanoma* Hs688(A).T	1.1	1.7	0.9	Bladder	1.2	2.6	1.7
Melanoma* Hs688(B).T	0.6	0.9	1.1	Gastric ca. (liver met.) NCI-N87	3.8	3.8	5.1
Melanoma* M14	1.4	0.8	1.7	Gastric ca. KATO.III	3.3	3.4	3.0
Melanoma* LOXIMVI	0.8	0.9	0.9	Colon ca. SW- 948	0.6	0.8	0.4
Melanoma* SK-MEL-5	4.9	4.1	3.8	Colon ca. SW480	3.9	5.1	4.9
Squamous cell carcinoma SCC-4	1.9	1.5	1.5	Colon ca.* (SW480 met) SW620	4.0	4.2	3.9
Testis Pool	0.7	0.7	0.9	Colon ca. HT29	1.4	0.8	1.3
Prostate ca.* (bone met) PC-3	3.5	3.7	3.4	Colon ca. HCT-116	4.2	5.0	4.9
Prostate Pool	1.2	1.1	1.1	Colon ca. CaCo-2	2.3	1.9	1.0
Placenta	0.6	0.4	0.8	Colon cancer tissue	2.0	2.9	2.6
Uterus Pool	0.1	0.4	0.3	Colon ca. SW1116	1.5	1.7	1.2
Ovarian ca. OVCAR-3	2.4	1.6	1.9	Colon ca. Colo-205	1.7	0.8	1.5
Ovarian ca. SK-OV-3	1.4	1.3	2.6	Colon ca. SW- 48	0.8	0.9	0.5
Ovarian ca. OVCAR-4	1.5	1.0	1.0	Colon Pool	1.7	1.8	1.7
Ovarian ca. OVCAR-5	10.0	6.6	7.9	Small Intestine Pool	4.3	3.3	4.1
Ovarian ca. IGROV-1	5.0	4.0	3.5	Stomach Pool	1.3	1.7	1.1
Ovarian ca. OVCAR-8	3.5	3.4	3.4	Bone Marrow Pool	0.8	0.7	0.7
Ovary	1.2	0.6	1.4	Fetal Heart	1.8	1.4	1.4
Breast ca. MCF-7	2.9	2.8	1.8	Heart Pool	4.7	5.0	5.2
Breast ca. MDA-MB- 231	3.8	5.0	6.0	Lymph Node Pool	3.4	3.0	1.8
Breast ca.	7.5	6.8	7.1	Fetal Skeletal	30.4	35.4	28.3

BT 549				Muscle			
Breast ca. T47D	14.3	19.8	21.3	Skeletal Muscle Pool	100.0	100.0	100.0
Breast ca. MDA-N	1.1	1.2	0.8	Spleen Pool	1.1	1.6	0.8
Breast Pool	1.6	2.1	1.6	Thymus Pool	2.3	3.2	3.5
Trachea	1.5	2.0	1.7	CNS cancer (glio/astro) U87-MG	3.4	4.7	4.8
Lung	0.4	0.4	0.8	CNS cancer (glio/astro) U-118-MG	3.7	3.7	5.3
Fetal Lung	3.1	3.2	4.1	CNS cancer (neuro;met) SK-N-AS	3.3	2.4	2.8
Lung ca. NCI-N417	0.8	0.6	1.3	CNS cancer (astro) SF-539	4.0	4.7	4.8
Lung ca. LX-1	5.3	3.4	3.8	CNS cancer (astro) SNB-75	15.8	14.5	17.4
Lung ca. NCI-H146	0.8	0.7	0.9	CNS cancer (glio) SNB-19	3.2	3.5	3.6
Lung ca. SHP-77	12.4	15.2	13.4	CNS cancer (glio) SF-295	7.9	10.4	8.3
Lung ca. A549	2.9	3.4	2.5	Brain (Amygdala) Pool	4.3	4.7	4.2
Lung ca. NCI-H526	1.1	1.1	0.9	Brain (cerebellum)	17.7	20.6	16.3
Lung ca. NCI-H23	10.2	9.6	10.4	Brain (fetal)	3.9	3.8	4.0
Lung ca. NCI-H460	2.1	1.6	0.9	Brain (Hippocampus) Pool	6.1	5.6	5.9
Lung ca. HOP-62	2.6	3.0	3.1	Cerebral Cortex Pool	5.2	4.8	4.8
Lung ca. NCI-H522	5.0	4.8	5.1	Brain (Substantia nigra) Pool	6.1	6.6	6.3
Liver	0.0	0.0	0.1	Brain (Thalamus) Pool	6.6	0.0	6.0
Fetal Liver	0.8	0.9	1.2	Brain (whole)	5.3	4.5	3.0
Liver ca. HepG2	1.5	0.7	1.2	Spinal Cord Pool	13.7	13.3	15.9
Kidney	5.8	6.3	5.7	Adrenal Gland	4.3	3.6	3.8

Pool							
Fetal Kidney	1.5	2.1	1.6	Pituitary gland Pool	1.0	0.7	0.7
Renal ca. 786-0	1.8	1.8	1.9	Salivary Gland	0.8	0.6	0.2
Renal ca. A498	1.2	0.9	1.0	Thyroid (female)	0.8	0.4	0.6
Renal ca. ACHN	4.8	4.1	4.1	Pancreatic ca. CAPAN2	3.8	4.4	5.2
Renal ca. UO-31	1.7	2.8	2.4	Pancreas Pool	2.8	3.5	2.0

Table BE. General\_screening\_panel\_v1.5

Tissue Name	Rel. Exp.(%) Ag5945, Run 247774858	Tissue Name	Rel. Exp.(%) Ag5945, Run 247774858
Adipose	1.6	Renal ca. TK-10	0.0
Melanoma* Hs688(A).T	0.3	Bladder	0.2
Melanoma* Hs688(B).T	0.0	Gastric ca. (liver met.) NCI-N87	0.0
Melanoma* M14	0.0	Gastric ca. KATO III	0.0
Melanoma* LOXIMVI	0.0	Colon ca. SW-948	0.0
Melanoma* SK-MEL-5	0.3	Colon ca. SW480	0.0
Squamous cell carcinoma SCC-4	0.0	Colon ca.* (SW480 met) SW620	0.0
Testis Pool	0.1	Colon ca. HT29	0.0
Prostate ca.* (bone met) PC-3	0.0	Colon ca. HCT-116	0.0
Prostate Pool	0.4	Colon ca. CaCo-2	0.0
Placenta	0.0	Colon cancer tissue	0.3
Uterus Pool	0.1	Colon ca. SW1116	0.0
Ovarian ca. OVCAR-3	0.0	Colon ca. Colo-205	0.0
Ovarian ca. SK-OV-3	0.0	Colon ca. SW-48	0.0
Ovarian ca. OVCAR-4	0.0	Colon Pool	0.2
Ovarian ca. OVCAR-5	0.0	Small Intestine Pool	0.6
Ovarian ca.	0.5	Stomach Pool	0.2

IGROV-1			
Ovarian ca. OVCAR-8	0.3	Bone Marrow Pool	0.1
Ovary	0.0	Fetal Heart	0.4
Breast ca. MCF-7	0.0	Heart Pool	2.8
Breast ca. MDA-MB-231	0.0	Lymph Node Pool	0.2
Breast ca. BT 549	0.6	Fetal Skeletal Muscle	16.0
Breast ca. T47D	0.0	Skeletal Muscle Pool	100.0
Breast ca. MDA-N	0.1	Spleen Pool	0.1
Breast Pool	0.2	Thymus Pool	0.1
Trachea	0.2	CNS cancer (glio/astro) U87-MG	0.0
Lung	0.0	CNS cancer (glio/astro) U-118-MG	0.6
Fetal Lung	0.4	CNS cancer (neuro;met) SK-N-AS	0.1
Lung ca. NCI-N417	0.0	CNS cancer (astro) SF-539	0.1
Lung ca. LX-1	0.0	CNS cancer (astro) SNB-75	2.1
Lung ca. NCI-H146	0.0	CNS cancer (glio) SNB-19	0.7
Lung ca. SHP-77	0.5	CNS cancer (glio) SF-295	0.0
Lung ca. A549	0.0	Brain (Amygdala) Pool	2.3
Lung ca. NCI-H526	0.0	Brain (cerebellum)	8.1
Lung ca. NCI-H23	0.0	Brain (fetal)	0.7
Lung ca. NCI-H460	0.0	Brain (Hippocampus) Pool	3.5
Lung ca. HOP-62	0.0	Cerebral Cortex Pool	2.0
Lung ca. NCI-H522	0.0	Brain (Substantia nigra) Pool	2.5
Liver	0.0	Brain (Thalamus) Pool	3.0
Fetal Liver	0.0	Brain (whole)	2.0
Liver ca. HepG2	0.0	Spinal Cord Pool	7.0
Kidney Pool	0.8	Adrenal Gland	1.0
Fetal Kidney	0.0	Pituitary gland Pool	0.3
Renal ca. 786-0	0.0	Salivary Gland	0.3
Renal ca. A498	0.1	Thyroid (female)	0.0
Renal ca. ACHN	0.0	Pancreatic ca. CAPAN2	0.0
Renal ca. UO-31	0.0	Pancreas Pool	0.3

Table BF. Panel 4.1D

Tissue Name	Rel. Exp.(%) Ag5945, Run 248173662	Tissue Name	Rel. Exp.(%) Ag5945, Run 248173662
Secondary Th1 act	0.0	HUVEC IL-1beta	0.0
Secondary Th2 act	0.0	HUVEC IFN gamma	0.0
Secondary Tr1 act	0.0	HUVEC TNF alpha + IFN gamma	0.0
Secondary Th1 rest	0.0	HUVEC TNF alpha + IL4	0.0
Secondary Th2 rest	0.0	HUVEC IL-11	1.3
Secondary Tr1 rest	0.0	Lung Microvascular EC none	0.0
Primary Th1 act	0.0	Lung Microvascular EC TNFalpha + IL-1beta	0.0
Primary Th2 act	0.0	Microvascular Dermal EC none	0.0
Primary Tr1 act	0.0	Microvascular Dermal EC TNFalpha + IL-1beta	0.0
Primary Th1 rest	0.0	Bronchial epithelium TNFalpha + IL1beta	0.0
Primary Th2 rest	0.0	Small airway epithelium none	0.0
Primary Tr1 rest	0.0	Small airway epithelium TNFalpha + IL-1beta	0.0
CD45RA CD4 lymphocyte act	0.0	Coronary artery SMC rest	0.0
CD45RO CD4 lymphocyte act	0.0	Coronary artery SMC TNFalpha + IL-1beta	0.0
CD8 lymphocyte act	0.0	Astrocytes rest	0.0
Secondary CD8 lymphocyte rest	0.0	Astrocytes TNFalpha + IL-1beta	0.0
Secondary CD8 lymphocyte act	0.0	KU-812 (Basophil) rest	2.6
CD4 lymphocyte none	0.0	KU-812 (Basophil) PMA/ionomycin	3.1
2ry Th1/Th2/Tr1_anti- CD95 CH11	0.0	CCD1106 (Keratinocytes) none	0.0
LAK cells rest	0.0	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	0.0
LAK cells IL-2	0.0	Liver cirrhosis	3.0
LAK cells IL-2+IL-12	0.0	NCI-H292 none	0.0
LAK cells IL-2+IFN gamma	0.0	NCI-H292 IL-4	0.0

LAK cells IL-2+ IL-18	0.0	NCI-H292 IL-9	0.0
LAK cells PMA/ionomycin	0.0	NCI-H292 IL-13	0.0
NK Cells IL-2 rest	0.0	NCI-H292 IFN gamma	0.0
Two Way MLR 3 day	0.0	HPAEC none	0.0
Two Way MLR 5 day	0.0	HPAEC TNF alpha + IL-1 beta	0.0
Two Way MLR 7 day	0.0	Lung fibroblast none	5.4
PBMC rest	0.0	Lung fibroblast TNF alpha + IL-1 beta	0.0
PBMC PWM	0.0	Lung fibroblast IL-4	3.0
PBMC PHA-L	0.0	Lung fibroblast IL-9	2.2
Ramos (B cell) none	0.0	Lung fibroblast IL-13	0.0
Ramos (B cell) ionomycin	0.0	Lung fibroblast IFN gamma	12.3
B lymphocytes PWM	0.0	Dermal fibroblast CCD1070 rest	0.0
B lymphocytes CD40L and IL-4	0.0	Dermal fibroblast CCD1070 TNF alpha	0.0
EOL-1 dbcAMP	0.0	Dermal fibroblast CCD1070 IL-1 beta	0.0
EOL-1 dbcAMP PMA/ionomycin	0.0	Dermal fibroblast IFN gamma	32.3
Dendritic cells none	0.0	Dermal fibroblast IL-4	15.8
Dendritic cells LPS	0.0	Dermal Fibroblasts rest	100.0
Dendritic cells anti- CD40	0.0	Neutrophils TNFa+LPS	0.0
Monocytes rest	0.0	Neutrophils rest	0.0
Monocytes LPS	0.0	Colon	0.0
Macrophages rest	0.0	Lung	6.0
Macrophages LPS	0.0	Thymus	2.2
HUVEC none	0.0	Kidney	2.5
HUVEC starved	0.0		

Table BG. Panel 5D

Tissue Name	Rel. Exp.(%) Ag3882, Run 170221179	Tissue Name	Rel. Exp.(%) Ag3882, Run 170221179
97457_Patient- 02go_adipose	1.4	94709_Donor 2 AM - A_adipose	0.2
97476_Patient- 07sk_skeletal muscle	7.4	94710_Donor 2 AM - B_adipose	0.8
97477 Patient-	0.7	94711_Donor 2 AM - C_adipose	0.5

07ut_uterus			
97478_Patient-07pl_placenta	0.8	94712_Donor 2 AD - A_adipose	4.4
97481_Patient-08sk_skeletal muscle	5.0	94713_Donor 2 AD - B_adipose	7.5
97482_Patient-08ut_uterus	0.0	94714_Donor 2 AD - C_adipose	6.2
97483_Patient-08pl_placenta	0.2	94742_Donor 3 U - A_Mesenchymal Stem Cells	0.9
97486_Patient-09sk_skeletal muscle	13.7	94743_Donor 3 U - B_Mesenchymal Stem Cells	0.0
97487_Patient-09ut_uterus	0.1	94730_Donor 3 AM - A_adipose	0.3
97488_Patient-09pl_placenta	0.8	94731_Donor 3 AM - B_adipose	0.6
97492_Patient-10ut_uterus	0.0	94732_Donor 3 AM - C_adipose	0.9
97493_Patient-10pl_placenta	1.4	94733_Donor 3 AD - A_adipose	4.1
97495_Patient-11go_adipose	1.1	94734_Donor 3 AD - B_adipose	0.2
97496_Patient-11sk_skeletal muscle	47.3	94735_Donor 3 AD - C_adipose	3.2
97497_Patient-11ut_uterus	0.3	77138_Liver_HepG2untreated	1.5
97498_Patient-11pl_placenta	0.6	73556_Heart_Cardiac stromal cells (primary)	0.0
97500_Patient-12go_adipose	1.7	81735_Small Intestine	5.4
97501_Patient-12sk_skeletal muscle	100.0	72409_Kidney_Proximal Convoluted Tubule	0.0
97502_Patient-12ut_uterus	0.6	82685_Small intestine_Duodenum	0.6
97503_Patient-12pl_placenta	0.1	90650_Adrenal_Adrenocortical adenoma	0.2
94721_Donor 2 U - A_Mesenchymal Stem Cells	0.8	72410_Kidney_HRCE	0.5
94722_Donor 2 U - B_Mesenchymal Stem Cells	0.5	72411_Kidney_HRE	0.0
94723_Donor 2 U - C_Mesenchymal Stem Cells	0.5	73139_Uterus_Uterine smooth muscle cells	1.0

**AI\_comprehensive\_panel\_v1.0 Summary:** Ag5945 Highest expression is seen in OA synovium (CT=29). In addition, moderate levels of expression are also seen in a cluster of samples from OA bone, synovium, and cartilage. Thus, expression of this gene could be used to differentiate between OA derived samples and other samples on this panel and as a marker of OA. Furthermore, therapeutic modulation of the expression or function of this gene may be useful in the treatment of OA.

**General\_screening\_panel\_v1.4 Summary:** Ag3882 Three experiments with the same probe and primer produce results that are in excellent agreement. Highest expression of this gene is seen in skeletal muscle (CTs=26-27). This gene is also expressed at moderate to low levels in pituitary, adipose, adrenal gland, pancreas, thyroid, fetal liver and adult and fetal skeletal muscle and heart. This widespread expression among these tissues suggests that this gene product may play a role in normal neuroendocrine and metabolic function and that dysregulated expression of this gene may contribute to neuroendocrine disorders or metabolic diseases, such as obesity and diabetes.

This gene is widely expressed in this panel, with moderate expression seen in brain, colon, gastric, lung, breast, ovarian, and melanoma cancer cell lines. This expression profile suggests a role for this gene product in cell survival and proliferation. Modulation of this gene product may be useful in the treatment of cancer.

This gene is also expressed at moderate to low levels in the CNS, including the hippocampus, thalamus, substantia nigra, amygdala, cerebellum and cerebral cortex. Therefore, therapeutic modulation of the expression or function of this gene may be useful in the treatment of neurologic disorders, such as Alzheimer's disease, Parkinson's disease, schizophrenia, multiple sclerosis, stroke and epilepsy.

**General\_screening\_panel\_v1.5 Summary:** Ag3882 Highest expression of this gene is seen in skeletal muscle (CT=24). Overall, expression of this gene is in agreement with Panel 1.4. Please see that panel for discussion of utility of this gene.

**Panel 4.1D Summary:** Ag5945 Expression is limited to dermal fibroblasts, with highest expression in resting dermal fibroblasts (CT=32.3). Thus, expression of this gene could be used to differentiate between resting and activated dermal fibroblasts. This expression also suggests that this gene may be involved in inflammatory conditions of the skin.

**Panel 5D Summary:** Ag5945 Moderate levels of expression are seen in skeletal muscle, while this gene is not expressed in the liver derived samples on adult liver or liver cell line samples on Panels 1.4 and 1.5 and this panel.

**C. CG102822-03: Glutamine synthase.**

5. Expression of gene CG102822-03 was assessed using the primer-probe sets Ag4225 and Ag5106, described in Tables CA and CB. Results of the RTQ-PCR runs are shown in Tables CC, CD, CE and CF.

Table CA. Probe Name Ag4225

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-cagaacaccttccaccatga-3'	20	104	468
Probe	TET-5'-ccacctcagcaagttcccacttaa-3'-TAMRA	26	124	469
Reverse	5'-tgaggcagggacatgtacac-3'	20	165	470

Table CB. Probe Name Ag5106

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-aggaatcagcatgggagatc-3'	20	749	471
Probe	TET-5'-ttgcatcgtgtgtgaagactttgg-3'-TAMRA	26	792	472
Reverse	5'-ggcttaggatcaaaggttgc-3'	20	825	473

- 10 Table CC. CNS\_neurodegeneration\_v1.0

Tissue Name	Rel. Exp.(%) Ag4225, Run 249266000	Rel. Exp.(%) Ag5106, Run 249286585	Tissue Name	Rel. Exp.(%) Ag4225, Run 249266000	Rel. Exp.(%) Ag5106, Run 249286585
AD 1 Hippo	10.3	9.6	Control (Path) 3 Temporal Ctx	12.5	12.0
AD 2 Hippo	17.4	17.9	Control (Path) 4 Temporal Ctx	22.8	22.2
AD 3 Hippo	4.0	3.6	AD 1 Occipital Ctx	11.0	14.2

AD 4 Hippo	4.6	4.8	AD 2 Occipital Ctx (Missing)	0.0	0.0
AD 5 Hippo	67.8	58.2	AD 3 Occipital Ctx	9.0	7.4
AD 6 Hippo	100.0	100.0	AD 4 Occipital Ctx	19.9	22.4
Control 2 Hippo	18.0	19.9	AD 5 Occipital Ctx	22.7	23.7
Control 4 Hippo	8.0	5.7	AD 6 Occipital Ctx	28.1	33.2
Control (Path) 3 Hippo	6.8	20.4	Control 1 Occipital Ctx	4.7	4.5
AD 1 Temporal Ctx	10.9	12.2	Control 2 Occipital Ctx	37.1	34.2
AD 2 Temporal Ctx	27.5	28.7	Control 3 Occipital Ctx	16.0	19.1
AD 3 Temporal Ctx	6.3	6.2	Control 4 Occipital Ctx	8.0	10.2
AD 4 Temporal Ctx	19.6	24.5	Control (Path) 1 Occipital Ctx	42.3	36.1
AD 5 Inf Temporal Ctx	66.4	69.3	Control (Path) 2 Occipital Ctx	8.1	6.6
AD 5 Sup. Temporal Ctx	36.3	33.7	Control (Path) 3 Occipital Ctx	6.9	5.8
AD 6 Inf Temporal Ctx	94.0	84.7	Control (Path) 4 Occipital Ctx	10.2	7.4
AD 6 Sup. Temporal Ctx	87.7	84.7	Control 1 Parietal Ctx	9.3	10.4

Control 1 Temporal Ctx	9.1	11.1	Control 2 Parietal Ctx	54.3	39.8
Control 2 Temporal Ctx	30.4	28.5	Control 3 Parietal Ctx	10.9	18.9
Control 3 Temporal Ctx	15.1	21.5	Control (Path) 1 Parietal Ctx	48.6	41.2
Control 3 Temporal Ctx	11.3	9.9	Control (Path) 2 Parietal Ctx	21.6	21.6
Control (Path) 1 Temporal Ctx	37.9	34.6	Control (Path) 3 Parietal Ctx	10.5	9.3
Control (Path) 2 Temporal Ctx	29.7	28.9	Control (Path) 4 Parietal Ctx	26.2	23.7

Table CD. General\_screening\_panel\_v1.5

Tissue Name	Rel. Exp.(%) Ag5106, Run 228727271	Tissue Name	Rel. Exp.(%) Ag5106, Run 228727271
Adipose	26.6	Renal ca. TK-10	12.1
Melanoma* Hs688(A).T	6.4	Bladder	27.0
Melanoma* Hs688(B).T	5.8	Gastric ca. (liver met.) NCI-N87	17.2
Melanoma* M14	7.5	Gastric ca. KATO III	2.4
Melanoma* LOXIMVI	0.2	Colon ca. SW-948	3.5
Melanoma* SK- MEL-5	6.9	Colon ca. SW480	11.3
Squamous cell carcinoma SCC-4	8.8	Colon ca.* (SW480 met) SW620	8.8
Testis Pool	15.6	Colon ca. HT29	8.1
Prostate ca.* (bone met) PC-3	8.8	Colon ca. HCT-116	11.6
Prostate Pool	7.1	Colon ca. CaCo-2	28.7
Placenta	22.5	Colon cancer tissue	13.2
Uterus Pool	9.4	Colon ca. SW1116	0.9
Ovarian ca.	11.3	Colon ca. Colo-205	0.3

OVCAR-3			
Ovarian ca. SK-OV-3	2.9	Colon ca. SW-48	3.0
Ovarian ca. OVCAR-4	7.6	Colon Pool	12.6
Ovarian ca. OVCAR-5	27.2	Small Intestine Pool	9.5
Ovarian ca. IGROV-1	6.7	Stomach Pool	13.8
Ovarian ca. OVCAR-8	3.1	Bone Marrow Pool	5.3
Ovary	13.8	Fetal Heart	11.0
Breast ca. MCF-7	4.4	Heart Pool	7.0
Breast ca. MDA-MB-231	8.0	Lymph Node Pool	11.7
Breast ca. BT 549	6.3	Fetal Skeletal Muscle	11.0
Breast ca. T47D	7.7	Skeletal Muscle Pool	61.1
Breast ca. MDA-N	3.3	Spleen Pool	10.8
Breast Pool	10.9	Thymus Pool	8.7
Trachea	38.2	CNS cancer (glio/astro) U87-MG	3.6
Lung	5.1	CNS cancer (glio/astro) U-118-MG	0.4
Fetal Lung	27.2	CNS cancer (neuro;met) SK-N-AS	7.1
Lung ca. NCI-N417	6.9	CNS cancer (astro) SF-539	14.4
Lung ca. LX-1	3.0	CNS cancer (astro) SNB-75	13.0
Lung ca. NCI-H146	5.1	CNS cancer (glio) SNB-19	6.8
Lung ca. SHP-77	5.8	CNS cancer (glio) SF-295	5.1
Lung ca. A549	3.3	Brain (Amygdala) Pool	26.8
Lung ca. NCI-H526	18.9	Brain (cerebellum)	100.0
Lung ca. NCI-H23	1.1	Brain (fetal)	13.2
Lung ca. NCI-H460	3.5	Brain (Hippocampus) Pool	36.6
Lung ca. HOP-62	4.1	Cerebral Cortex Pool	64.2
Lung ca. NCI-H522	1.0	Brain (Substantia nigra) Pool	45.7
Liver	7.2	Brain (Thalamus) Pool	55.9
Fetal Liver	31.0	Brain (whole)	55.9
Liver ca. HepG2	23.7	Spinal Cord Pool	32.8

Kidney Pool	16.6	Adrenal Gland	11.3
Fetal Kidney	4.9	Pituitary gland Pool	2.6
Renal ca. 786-0	0.0	Salivary Gland	5.5
Renal ca. A498	0.0	Thyroid (female)	12.2
Renal ca. ACHN	4.2	Pancreatic ca. CAPAN2	5.1
Renal ca. UO-31	3.5	Pancreas Pool	12.8

Table CE. Panel 5 Islet

Tissue Name	Rel. Exp.(%) Ag4225, Run 248989150	Rel. Exp.(%) Ag4225, Run 2492529 11	Rel. Exp.(%) Ag5106, Run 31285250 4	Tissue Name	Rel. Exp.(%) Ag4225, Run 2489891 50	Rel. Exp.(%) Ag4225, Run 24925291 1	Rel. Exp.(%) Ag5106, Run 31285250 4
97457_Patient- 02go_adipose	36.3	48.6	42.0	94709_Donor 2 AM - A_adipose	15.6	27.9	15.4
97476_Patient- 07sk_skeletal muscle	16.7	17.4	0.0	94710_Donor 2 AM - B_adipose	10.6	18.9	15.3
97477_Patient- 07ut_uterus	12.0	15.9	10.6	94711_Donor 2 AM - C_adipose	7.4	14.5	12.5
97478_Patient- 07pl_placenta	15.4	27.4	23.3	94712_Donor 2 AD - A_adipose	17.1	22.1	34.9
99167_Bayer Patient 1	37.4	29.9	20.0	94713_Donor 2 AD - B_adipose	15.9	27.9	45.4
97482_Patient- 08ut_uterus	9.0	12.7	7.3	94714_Donor 2 AD - C_adipose	16.0	25.5	29.5
97483_Patient- 08pl_placenta	12.0	17.6	14.7	94742_Donor 3 U - A_Mesenchym al Stem Cells	1.8	3.8	2.3
97486_Patient- 09sk_skeletal muscle	7.6	9.3	9.4	94743_Donor 3 U - B_Mesenchyma l Stem Cells	4.3	4.6	2.5
97487_Patient- 09ut_uterus	19.5	21.0	11.2	94730_Donor 3 AM - A_adipose	15.0	20.2	28.5
97488_Patient- 09pl_placenta	9.6	22.2	13.8	94731_Donor 3 AM - B_adipose	9.9	13.7	46.0

97492_Patient-10ut_uterus	15.8	20.6	13.3	94732_Donor 3 AM - C_adipose	8.8	17.1	31.9
97493_Patient-10pl_placenta	43.2	52.5	38.4	94733_Donor 3 AD - A_adipose	6.7	6.7	14.1
97495_Patient-11go_adipose	33.4	33.9	18.8	94734_Donor 3 AD - B_adipose	2.2	4.7	11.4
97496_Patient-11sk_skeletal muscle	35.6	52.1	27.7	94735_Donor 3 AD - C_adipose	4.4	4.6	3.7
97497_Patient-11ut_uterus	18.9	22.8	19.9	77138_Liver_H epG2untreated	70.2	98.6	100.0
97498_Patient-11pl_placenta	17.1	19.1	9.0	73556_Heart_C ardiac stromal cells (primary)	3.6	4.4	3.1
97500_Patient-12go_adipose	100.0	100.0	73.2	81735_Small Intestine	21.6	19.9	16.4
97501_Patient-12sk_skeletal muscle	63.7	74.2	59.5	72409_Kidney_ Proximal Convolutated Tubule	2.0	2.2	7.7
97502_Patient-12ut_uterus	16.6	17.6	17.1	82685_Small intestine_Duod enum	6.6	10.8	7.4
97503_Patient-12pl_placenta	25.2	35.6	35.8	90650_Adrenal _Adrenocortical adenoma	6.6	8.1	5.1
94721_Donor 2 U - A_Mesenchyma l Stem Cells	4.5	7.5	10.3	72410_Kidney_ HRCE	13.1	10.4	7.6
94722_Donor 2 U - B_Mesenchyma l Stem Cells	4.2	5.6	5.2	72411_Kidney_ HRE	7.5	9.1	5.2
94723_Donor 2 U - C_Mesenchyma l Stem Cells	5.6	1.1	8.5	73139_Uterus_ Uterine smooth muscle cells	2.7	4.5	8.2

Table CF. Panel 5D

Tissue Name	Rel. Exp.(%) Ag4225, Run 181457566	Tissue Name	Rel. Exp.(%) Ag4225, Run 181457566
-------------	--	-------------	--

97457_Patient-02go_adipose	52.1	94709_Donor 2 AM - A_adipose	24.3
97476_Patient-07sk_skeletal muscle	16.4	94710_Donor 2 AM - B_adipose	15.8
97477_Patient-07ut_uterus	13.8	94711_Donor 2 AM - C_adipose	11.7
97478_Patient-07pl_placenta	24.5	94712_Donor 2 AD - A_adipose	22.1
97481_Patient-08sk_skeletal muscle	13.3	94713_Donor 2 AD - B_adipose	25.2
97482_Patient-08ut_uterus	12.0	94714_Donor 2 AD - C_adipose	23.5
97483_Patient-08pl_placenta	17.3	94742_Donor 3 U - A Mesenchymal Stem Cells	4.1
97486_Patient-09sk_skeletal muscle	9.2	94743_Donor 3 U - B Mesenchymal Stem Cells	5.5
97487_Patient-09ut_uterus	21.6	94730_Donor 3 AM - A_adipose	26.1
97488_Patient-09pl_placenta	21.3	94731_Donor 3 AM - B_adipose	12.9
97492_Patient-10ut_uterus	16.6	94732_Donor 3 AM - C_adipose	13.0
97493_Patient-10pl_placenta	52.5	94733_Donor 3 AD - A_adipose	8.4
97495_Patient-11go_adipose	39.5	94734_Donor 3 AD - B_adipose	4.9
97496_Patient-11sk_skeletal muscle	51.4	94735_Donor 3 AD - C_adipose	5.4
97497_Patient-11ut_uterus	24.8	77138_Liver_HepG2untreated	100.0
97498_Patient-11pl_placenta	23.2	73556_Heart_Cardiac stromal cells (primary)	3.5
97500_Patient-12go_adipose	92.7	81735_Small Intestine	19.5
97501_Patient-12sk_skeletal muscle	72.7	72409_Kidney_Proximal Convoluted Tubule	2.3
97502_Patient-12ut_uterus	26.2	82685_Small intestine_Duodenum	10.0
97503_Patient-12pl_placenta	27.0	90650_Adrenal_Adrenocortical adenoma	6.4
94721_Donor 2 U - A_Mesenchymal Stem Cells	5.4	72410_Kidney_HRCE	10.3
94722_Donor 2 U - B_Mesenchymal	5.6	72411_Kidney_HRE	8.0

Stem Cells			
94723_Donor 2 U - C_Mesenchymal Stem Cells	6.4	73139_Uterus_Uterine smooth muscle cells	3.7

**CNS\_neurodegeneration\_v1.0 Summary:** Ag4225/Ag5106 Two experiments with two different probe and primer sets produce results that are in excellent agreement, with highest expression in the hippocampus of an Alzheimer's patient (CTs=23-24). This panel does not show differential expression of this gene in Alzheimer's disease. However, this profile confirms the expression of this gene at moderate levels in the brain. Please see Panel 1.5 for discussion of utility of this gene in the central nervous system.

**General\_screening\_panel\_v1.4 Summary:** Ag4225 Results from one experiment with this gene are not included. The amp plot indicates that there were experimental difficulties with this run.

**General\_screening\_panel\_v1.5 Summary:** Ag5106 Expression of this gene appears to have a brain-preferential distribution among normal tissues, with highest expression seen in the cerebellum (CT=22). This gene is also expressed at high levels throughout the CNS, including the hippocampus, thalamus, substantia nigra, amygdala, cerebellum and cerebral cortex. Therefore, therapeutic modulation of the expression or function of this gene may be useful in the treatment of neurological disorders, such as Alzheimer's disease, Parkinson's disease, schizophrenia, multiple sclerosis, stroke and epilepsy.

Among tissues with metabolic function, this gene is expressed at high levels in pituitary, adipose, adrenal gland, pancreas, thyroid, and adult and fetal skeletal muscle, heart, and liver. This widespread expression among these tissues suggests that this gene product may play a role in normal neuroendocrine and metabolic function and that dysregulated expression of this gene may contribute to neuroendocrine disorders or metabolic diseases, such as obesity and diabetes.

**Panel 5 Islet Summary:** Ag4225/Ag5106 Multiple experiments with two different probe and primer sets produce results that are in excellent agreement, with highest expression in a liver cell line and adipose from a diabetic patient (CTs=26.5). In addition, high to moderate levels of expression are seen in metabolic tissues, including placenta, adipose and skeletal muscle, in agreement with Panel 1.5. This gene encodes glutamine synthase (GS) and also

- appears to be slightly up-regulated in diabetic skeletal muscle (patient 12). Up-regulation of glutamine synthase, which is critical for glutamine production, has been reported in obesity and diabetes, as well as in some myopathies. Muscle catabolism leads to the release of glutamine and contributes to gluconeogenesis in the liver. Inhibition of GS may decrease glutamine production, inhibit gluconeogenesis and necessitate fatty acid oxidation for energy generation. Therefore, an antagonist of glutamine synthase may be beneficial in treatment of obesity and diabetes.

- Panel 5D Summary:** Ag4225 Highest expression is in a liver cell line (CT=26.6). Expression is in agreement with Panel 5I. Please see that panel for further discussion of expression and utility of this gene in obesity and diabetes.

**D. CG103241-02: UDPGAL:GLCNAC B1,4 GALACTOSYLTRANSFERASE.**

Expression of gene CG103241-02 was assessed using the primer-probe set Ag7620, described in Table DA.

Table DA. Probe Name Ag7620

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5' -ctgagtaaggctcagtttctgaga-3'	24	830	474
Probe	TET-5' -tcaatggcttccccaatgagtactgg-3' - TAMRA	26	855	475
Reverse	5' -aatcttggttaaaccggttgaag-3'	22	907	476

- CNS\_neurodegeneration\_v1.0 Summary:** Ag7620 Expression of this gene is low/undetectable in all samples on this panel (CTs>35). (Data not shown).

**Panel 4.1D Summary:** Ag7620 Expression of this gene is low/undetectable in all samples on this panel (CTs>35). (Data not shown).

**E. CG106249-02: Kinesin.**

- Expression of gene CG106249-02 was assessed using the primer-probe set Ag7282, described in Table EA. Results of the RTQ-PCR runs are shown in Tables EB and EC.

Table EA. Probe Name Ag7282

Primers	Sequences	Length	Start	SEQ ID
---------	-----------	--------	-------	--------

			<b>Position</b>	<b>No</b>
<b>Forward</b>	5'-atcccaaagaaggcccttat-3'	20	550	477
<b>Probe</b>	TET-5'-cgtcaccataattctgtactaaatgtttgg-3'-TAMRA	30	583	478
<b>Reverse</b>	5'-cccgcatccataagttcttc-3'	20	615	479

Table EB. CNS\_neurodegeneration\_v1.0

<b>Tissue Name</b>	<b>Rel. Exp.(%) Ag7282, Run 296560376</b>	<b>Tissue Name</b>	<b>Rel. Exp.(%) Ag7282, Run 296560376</b>
AD 1 Hippo	12.5	Control (Path) 3 Temporal Ctx	15.5
AD 2 Hippo	25.3	Control (Path) 4 Temporal Ctx	28.3
AD 3 Hippo	13.7	AD 1 Occipital Ctx	27.4
AD 4 Hippo	11.7	AD 2 Occipital Ctx (Missing)	0.0
AD 5 Hippo	100.0	AD 3 Occipital Ctx	8.5
AD 6 Hippo	59.5	AD 4 Occipital Ctx	11.0
Control 2 Hippo	38.7	AD 5 Occipital Ctx	33.2
Control 4 Hippo	19.1	AD 6 Occipital Ctx	15.7
Control (Path) 3 Hippo	12.9	Control 1 Occipital Ctx	7.7
AD 1 Temporal Ctx	42.0	Control 2 Occipital Ctx	48.0
AD 2 Temporal Ctx	12.7	Control 3 Occipital Ctx	38.7
AD 3 Temporal Ctx	10.2	Control 4 Occipital Ctx	10.5
AD 4 Temporal Ctx	35.6	Control (Path) 1. Occipital Ctx	57.8
AD 5 Inf Temporal Ctx	94.0	Control (Path) 2 Occipital Ctx	13.1
AD 5 Sup Temporal Ctx	57.8	Control (Path) 3 Occipital Ctx	7.0
AD 6 Inf Temporal Ctx	33.2	Control (Path) 4. Occipital Ctx	19.1
AD 6 Sup Temporal Ctx	48.6	Control 1 Parietal Ctx	12.7
Control 1 Temporal Ctx	10.7	Control 2 Parietal Ctx	53.6
Control 2 Temporal Ctx	15.1	Control 3 Parietal Ctx	21.0
Control 3 Temporal Ctx	32.1	Control (Path) 1 Parietal Ctx	61.1

Control 3 Temporal Ctx	6.4	Control (Path) 2 Parietal Ctx	28.7
Control (Path) 1 Temporal Ctx	45.7	Control (Path) 3 Parietal Ctx	9.7
Control (Path) 2 Temporal Ctx	51.1	Control (Path) 4 Parietal Ctx	31.9

Table EC. Panel 4.1D

Tissue Name	Rel. Exp.(%) Ag7282, Run 296559398	Tissue Name	Rel. Exp.(%) Ag7282, Run 296559398
Secondary Th1 act	33.2	HUVEC IL-1beta	12.6
Secondary Th2 act	35.8	HUVEC IFN gamma	20.3
Secondary Tr1 act	8.8	HUVEC TNF alpha + IFN gamma	3.1
Secondary Th1 rest	2.5	HUVEC TNF alpha + IL4	0.0
Secondary Th2 rest	3.4	HUVEC IL-11	14.6
Secondary Tr1 rest	3.0	Lung Microvascular EC none	22.1
Primary Th1 act	0.0	Lung Microvascular EC TNFalpha + IL-1beta	6.5
Primary Th2 act	7.5	Microvascular Dermal EC none	3.3
Primary Tr1 act	10.6	Microvascular Dermal EC TNFalpha + IL-1beta	0.0
Primary Th1 rest	2.0	Bronchial epithelium TNFalpha + IL1beta	18.7
Primary Th2 rest	0.0	Small airway epithelium none	24.8
Primary Tr1 rest	0.0	Small airway epithelium TNFalpha + IL-1beta	49.0
CD45RA CD4 lymphocyte act	12.8	Coronary artery SMC rest	9.8
CD45RO CD4 lymphocyte act	46.0	Coronary artery SMC TNFalpha + IL-1beta	9.6
CD8 lymphocyte act	12.2	Astrocytes rest	0.0
Secondary CD8 lymphocyte rest	5.3	Astrocytes TNFalpha + IL-1beta	3.5
Secondary CD8 lymphocyte act	0.0	KU-812 (Basophil) rest	38.7
CD4 lymphocyte none	6.0	KU-812 (Basophil) PMA/ionomycin	48.6
2ry Th1/Th2/Tr1_anti- CD95 CH11	5.0	CCD1106 (Keratinocytes) none	39.8

LAK cells rest	9.5	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	9.0
LAK cells IL-2	6.6	Liver cirrhosis	12.5
LAK cells IL-2+IL-12	0.0	NCI-H292 none	12.5
LAK cells IL-2+IFN gamma	6.8	NCI-H292 IL-4	13.9
LAK cells IL-2+ IL-18	4.5	NCI-H292 IL-9	26.6
LAK cells PMA/ionomycin	3.7	NCI-H292 IL-13	16.7
NK Cells IL-2 rest	22.8	NCI-H292 IFN gamma	2.1
Two Way MLR 3 day	8.2	HPAEC none	5.1
Two Way MLR 5 day	3.3	HPAEC TNF alpha + IL- 1 beta	13.8
Two Way MLR 7 day	0.0	Lung fibroblast none	26.8
PBMC rest	2.4	Lung fibroblast TNF alpha + IL-1 beta	17.0
PBMC PWM	2.4	Lung fibroblast IL-4	11.1
PBMC PHA-L	8.1	Lung fibroblast IL-9	8.7
Ramos (B cell) none	10.1	Lung fibroblast IL-13	7.7
Ramos (B cell) ionomycin	13.0	Lung fibroblast IFN gamma	20.6
B lymphocytes PWM	7.4	Dermal fibroblast CCD1070 rest	6.9
B lymphocytes CD40L and IL-4	18.2	Dermal fibroblast CCD1070 TNF alpha	6.3
EOL-1 dbcAMP	16.4	Dermal fibroblast CCD1070 IL-1 beta	0.0
EOL-1 dbcAMP PMA/ionomycin	4.7	Dermal fibroblast IFN gamma	10.2
Dendritic cells none	7.3	Dermal fibroblast IL-4	26.2
Dendritic cells LPS	3.0	Dermal Fibroblasts rest	24.5
Dendritic cells anti- CD40	8.2	Neutrophils TNFa+LPS	0.0
Monocytes rest	3.8	Neutrophils rest	4.6
Monocytes LPS	11.6	Colon	4.8
Macrophages rest	12.5	Lung	2.5
Macrophages LPS	6.0	Thymus	12.5
HUVEC none	6.3	Kidney	100.0
HUVEC starved	18.3		

**CNS\_neurodegeneration\_v1.0 Summary:** Ag7282 This panel confirms the expression of this gene at very low levels in the brains of an independent group of individuals. No

differential expression of this gene was detected between Alzheimer's diseased postmortem brains and those of non-demented controls in this experiment. However, this panel confirms the expression of this gene at very low levels in the brains of an independent group of individuals. Therefore, therapeutic modulation of this gene product may be useful in the treatment of central nervous system disorders such as Parkinson's disease, epilepsy, multiple sclerosis, schizophrenia and depression.

**Panel 4.1D Summary:** Ag7282 Low levels of expression of this gene is seen mainly in kidney (CT=34.3). Therefore, expression of this gene may be used to distinguish kidney from other samples used in this panel. In addition, therapeutic targeting of the expression or function of this gene may modulate kidney function and be important in the treatment of inflammatory or autoimmune diseases that affect the kidney, including lupus and glomerulonephritis.

#### **F. CG119418-01: farnesyl-diphosphate farnesyltransferase 1.**

Expression of gene CG119418-01 was assessed using the primer-probe set Ag4508, described in Table FA. Results of the RTQ-PCR runs are shown in Tables FB and FC.

Table FA. Probe Name Ag4508

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-gaagaccccttagttggtgaag-3'	22	586	480
Probe	TET-5'-caactctatgggcctgtttctgcaga-3'-TAMRA	26	621	481
Reverse	5'-ccagatagtcacggatgatgtt-3'	22	652	482

Table FB. General\_screening\_panel\_v1.4

Tissue Name	Rel. Exp.(%) Ag4508, Run 213805830	Tissue Name	Rel. Exp.(%) Ag4508, Run 213805830
Adipose	4.5	Renal ca. TK-10	23.2
Melanoma* Hs688(A).T	9.2	Bladder	8.8
Melanoma* Hs688(B).T	11.9	Gastric ca. (liver met.) NCI-N87	28.5
Melanoma*. M14	30.1	Gastric ca. KATO III	75.3
Melanoma* LOXIMVI	14.8	Colon ca. SW-948	16.0

Melanoma* SK-MEL-5	25.5	Colon ca. SW480	18.3
Squamous cell carcinoma SCC-4	17.4	Colon ca.* (SW480 met) SW620	18.0
Testis Pool	10.2	Colon ca. HT29	17.2
Prostate ca.* (bone met) PC-3	5.3	Colon ca. HCT-116	32.1
Prostate Pool	5.2	Colon ca. CaCo-2	33.7
Placenta	5.0	Colon cancer tissue	8.7
Uterus Pool	2.7	Colon ca. SW1116	3.8
Ovarian ca. OVCAR-3	17.7	Colon ca. Colo-205	13.2
Ovarian ca. SK-OV-3	25.9	Colon ca. SW-48	11.9
Ovarian ca. OVCAR-4	12.4	Colon Pool	5.3
Ovarian ca. OVCAR-5	22.2	Small Intestine Pool	6.0
Ovarian ca. IGROV-1	19.1	Stomach Pool	3.3
Ovarian ca. OVCAR-8	4.6	Bone Marrow Pool	2.7
Ovary	8.0	Fetal Heart	2.7
Breast ca. MCF-7	15.8	Heart Pool	3.3
Breast ca. MDA-MB-231	14.0	Lymph Node Pool	6.3
Breast ca. BT 549	100.0	Fetal Skeletal Muscle	2.8
Breast ca. T47D	48.3	Skeletal Muscle Pool	6.9
Breast ca. MDA-N	18.0	Spleen Pool	3.0
Breast Pool	5.1	Thymus Pool	4.0
Trachea	9.2	CNS cancer (glio/astro) U87-MG	18.4
Lung	1.9	CNS cancer (glio/astro) U-118-MG	9.4
Fetal Lung	10.2	CNS cancer (neuro;met) SK-N-AS	18.3
Lung ca. NCI-N417	9.2	CNS cancer (astro) SF-539	55.5
Lung ca. LX-1	27.5	CNS cancer (astro) SNB-75	20.4
Lung ca. NCI-H146	15.2	CNS cancer (glio) SNB-19	16.5
Lung ca. SHP-77	35.4	CNS cancer (glio) SF-295	15.9

Lung ca. A549	20.7	Brain (Amygdala) Pool	7.3
Lung ca. NCI-H526	8.4	Brain (cerebellum)	10.1
Lung ca. NCI-H23	8.8	Brain (fetal)	22.1
Lung ca. NCI-H460	6.0	Brain (Hippocampus) Pool	8.1
Lung ca. HOP-62	13.1	Cerebral Cortex Pool	8.9
Lung ca. NCI-H522	8.0	Brain (Substantia nigra) Pool	7.5
Liver	1.8	Brain (Thalamus) Pool	11.3
Fetal Liver	33.7	Brain (whole)	12.9
Liver ca. HepG2	36.3	Spinal Cord Pool	11.3
Kidney Pool	8.7	Adrenal Gland	15.5
Fetal Kidney	4.6	Pituitary gland Pool	2.1
Renal ca. 786-0	14.6	Salivary Gland	7.6
Renal ca. A498	2.0	Thyroid (female)	3.9
Renal ca. ACHN	27.4	Pancreatic ca. CAPAN2	36.9
Renal ca. UO-31	18.6	Pancreas Pool	5.4

Table FC. Panel 5 Islet

Tissue Name	Rel. Exp.(%) Ag4508, Run 200923967	Tissue Name	Rel. Exp.(%) Ag4508, Run 200923967
97457_Patient-02go_adipose	7.7	94709_Donor 2 AM - A_adipose	9.8
97476_Patient-07sk_skeletal muscle	7.4	94710_Donor 2 AM - B_adipose	7.7
97477_Patient-07ut_uterus	4.5	94711_Donor 2 AM - C_adipose	5.5
97478_Patient-07pl_placenta	12.4	94712_Donor 2 AD - A_adipose	14.6
99167_Bayer Patient 1	30.8	94713_Donor 2 AD - B_adipose	18.8
97482_Patient-08ut_uterus	3.4	94714_Donor 2 AD - C_adipose	16.5
97483_Patient-08pl_placenta	13.3	94742_Donor 3 U - A_Mesenchymal Stem Cells	5.7
97486_Patient-09sk_skeletal muscle	5.5	94743_Donor 3 U - B_Mesenchymal Stem Cells	9.0
97487_Patient-09ut_uterus	7.7	94730_Donor 3 AM - A_adipose	10.1
97488_Patient-09pl_placenta	7.0	94731_Donor 3 AM - B_adipose	5.7

97492_Patient-10ut_uterus	8.0	94732_Donor 3 AM - C_adipose	7.1
97493_Patient-10pl_placenta	23.8	94733_Donor 3 AD - A_adipose	20.3
97495_Patient-11go_adipose	7.1	94734_Donor 3 AD - B_adipose	6.7
97496_Patient-11sk_skeletal muscle	16.5	94735_Donor 3 AD - C_adipose	16.2
97497_Patient-11ut_uterus	9.6	77138_Liver_HepG2untreated	100.0
97498_Patient-11pl_placenta	7.5	73556_Heart_Cardiac stromal cells (primary)	11.5
97500_Patient-12go_adipose	13.0	81735_Small Intestine	21.6
97501_Patient-12sk_skeletal muscle	47.3	72409_Kidney Proximal Convoluted Tubule	20.9
97502_Patient-12ut_uterus	8.8	82685_Small intestine Duodenum	7.0
97503_Patient-12pl_placenta	13.0	90650_Adrenal_Adrenocortical adenoma	5.4
94721_Donor 2 U - A_Mesenchymal Stem Cells	17.6	72410_Kidney_HRCE	58.6
94722_Donor 2 U - B_Mesenchymal Stem Cells	8.8	72411_Kidney_HRE	50.0
94723_Donor 2 U - C_Mesenchymal Stem Cells	11.4	73139_Uterus_Uterine smooth muscle cells	20.0

**General\_screening\_panel\_v1.4 Summary:** Ag4508 Highest expression of this gene is detected in a breast cancer BT 549 cell line (CT=23.6). High expression of this gene is also seen in cluster of cancer cell lines derived from pancreatic, gastric, colon, lung, liver, renal, breast, ovarian, prostate, squamous cell carcinoma, melanoma and brain cancers. Thus, expression of this gene could be used as a marker to detect the presence of these cancers. Furthermore, therapeutic modulation of the expression or function of this gene may be effective in the treatment of pancreatic, gastric, colon, lung, liver, renal, breast, ovarian, prostate, squamous cell carcinoma, melanoma and brain cancers.

Among tissues with metabolic or endocrine function, this gene is expressed at high levels in pancreas, adipose, adrenal gland, thyroid, pituitary gland, skeletal muscle, heart, liver and the gastrointestinal tract. Therefore, therapeutic modulation of the activity of this gene

may prove useful in the treatment of endocrine/metabolically related diseases, such as obesity and diabetes.

In addition, this gene is expressed at high levels in all regions of the central nervous system examined, including amygdala, hippocampus, substantia nigra, thalamus, cerebellum, cerebral cortex, and spinal cord. Therefore, therapeutic modulation of this gene product may be useful in the treatment of central nervous system disorders such as Alzheimer's disease, Parkinson's disease, epilepsy, multiple sclerosis, schizophrenia and depression.

Interestingly, this gene is expressed at much higher levels in fetal (CT=25) when compared to adult liver (CT=29). This observation suggests that expression of this gene can be used to distinguish fetal from adult liver. In addition, the relative overexpression of this gene in fetal tissue suggests that the protein product may enhance liver growth or development in the fetus and thus may also act in a regenerative capacity in the adult. Therefore, therapeutic modulation of the protein encoded by this gene could be useful in treatment of liver related diseases.

**Panel 5 Islet Summary:** Ag4508 Highest expression of this gene is detected in liver cancer HepG2 cell line (CT=25.3). This gene shows a wide spread expression in this panel, which correlates with the expression in panel 1.4. High expression of this gene is detected in islet cells, adipose, skeletal muscle, uterus, placenta, heart smooth muscle, small intestine and kidney. This gene codes for Farnesyl-diphosphate farnesyltransferase. Farnesyl-diphosphate farnesyltransferase is involved in the cholesterol biosynthetic pathway. The operation of this pathway appears to be important for glucose homeostasis and insulin secretion in pancreatic beta cells (Flamez D, Berger V, Kruhoffer M, Orntoft T, Pipeleers D, Schuit FC., 2002, Critical role for cataplerosis via citrate in glucose-regulated insulin release. Diabetes. 2002 Jul;51(7):2018-24. PMID: 12086928). Therefore, therapeutic modulation of this gene product may enhance insulin secretion in Type 2 diabetes.

#### **G. CG120359-01: acetyl-CoA synthetase.**

Expression of gene CG120359-01 was assessed using the primer-probe set Ag4830, described in Table GA. Results of the RTQ-PCR runs are shown in Tables GB and GC.

Table GA. Probe Name Ag4830

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-gtggagcattgtggacaaatac-3'	22	1182	483
Probe	TET-5'-tgaccaagttctacacagcaccaca-3'- TAMRA	26	1208	484
Reverse	5'-gctcatctccaaacttcagag-3'	22	1246	485

Table GB. General\_screening\_panel\_v1.4

Tissue Name	Rel. Exp.(%) Ag4830, Run 213856337	Tissue Name	Rel. Exp.(%) Ag4830, Run 213856337
Adipose	16.2	Renal ca. TK-10	39.8
Melanoma* Hs688(A).T	13.1	Bladder	20.9
Melanoma* Hs688(B).T	12.6	Gastric ca. (liver met.) NCI-N87	36.6
Melanoma* M14	47.6	Gastric ca. KATO III	37.6
Melanoma* LOXIMVI	7.4	Colon ca. SW-948	12.8
Melanoma* SK- MEL-5	21.6	Colon ca. SW480	88.9
Squamous cell carcinoma SCC-4	17.3	Colon ca.* (SW480 met) SW620	27.2
Testis Pool	9.2	Colon ca. HT29	9.9
Prostate ca.* (bone met) PC-3	59.9	Colon ca. HCT-116	24.7
Prostate Pool	6.6	Colon ca. CaCo-2	62.9
Placenta	16.6	Colon cancer tissue	32.8
Uterus Pool	5.0	Colon ca. SW1116	6.0
Ovarian ca. OVCAR-3	22.2	Colon ca. Colo-205	7.7
Ovarian ca. SK- OV-3	13.8	Colon ca. SW-48	48.6
Ovarian ca. OVCAR-4	22.4	Colon Pool	10.9
Ovarian ca. OVCAR-5	45.4	Small Intestine Pool	12.6
Ovarian ca. IGROV-1	56.6	Stomach Pool	7.2
Ovarian ca. OVCAR-8	9.7	Bone Marrow Pool	4.8
Ovary	8.5	Fetal Heart	11.8
Breast ca. MCF-7	9.7	Heart Pool	13.1

Breast ca. MDA-MB-231	32.8	Lymph Node Pool	12.0
Breast ca. BT 549	28.3	Fetal Skeletal Muscle	20.3
Breast ca. T47D	88.3	Skeletal Muscle Pool	44.4
Breast ca. MDA-N	34.4	Spleen Pool	5.8
Breast Pool	9.3	Thymus Pool	10.3
Trachea	12.2	CNS cancer (glio/astro) U87-MG	49.3
Lung	4.0	CNS cancer (glio/astro) U-118-MG	24.3
Fetal Lung	27.5	CNS cancer (neuro;met) SK-N-AS	24.0
Lung ca. NCI-N417	1.6	CNS cancer (astro) SF-539	14.5
Lung ca. LX-1	26.2	CNS cancer (astro) SNB-75	33.9
Lung ca. NCI-H146	1.6	CNS cancer (glio) SNB-19	51.4
Lung ca. SHP-77	6.8	CNS cancer (glio) SF-295	30.8
Lung ca. A549	13.7	Brain (Amygdala) Pool	9.5
Lung ca. NCI-H526	2.1	Brain (cerebellum)	21.3
Lung ca. NCI-H23	19.6	Brain (fetal)	11.0
Lung ca. NCI-H460	13.3	Brain (Hippocampus) Pool	7.3
Lung ca. HOP-62	19.2	Cerebral Cortex Pool	10.3
Lung ca. NCI-H522	11.7	Brain (Substantia nigra) Pool	12.9
Liver	5.8	Brain (Thalamus) Pool	10.8
Fetal Liver	65.5	Brain (whole)	10.6
Liver ca. HepG2	55.5	Spinal Cord Pool	8.8
Kidney Pool	15.4	Adrenal Gland	62.4
Fetal Kidney	5.7	Pituitary gland Pool	1.6
Renal ca. 786-0	13.6	Salivary Gland	13.4
Renal ca. A498	8.4	Thyroid (female)	5.8
Renal ca. ACHN	100.0	Pancreatic ca. CAPAN2	56.6
Renal ca. UO-31	18.6	Pancreas Pool	11.6

Table GC. Panel 5 Islet

Tissue Name	Rel. Exp.(%) Ag4830, Run 223846062	Tissue Name	Rel. Exp.(%) Ag4830, Run 223846062
-------------	--	-------------	--

97457_Patient-02go_adipose	27.9	94709_Donor 2 AM - A_adipose	10.1
97476_Patient-07sk_skeletal muscle	19.2	94710_Donor 2 AM - B_adipose	11.4
97477_Patient-07ut_uterus	5.2	94711_Donor 2 AM - C_adipose	0.6
97478_Patient-07pl_placenta	15.7	94712_Donor 2 AD - A_adipose	5.3
99167_Bayer Patient 1	43.8	94713_Donor 2 AD - B_adipose	10.3
97482_Patient-08ut_uterus	1.1	94714_Donor 2 AD - C_adipose	10.4
97483_Patient-08pl_placenta	12.5	94742_Donor 3 U - A_Mesenchymal Stem Cells	1.4
97486_Patient-09sk_skeletal muscle	11.5	94743_Donor 3 U - B_Mesenchymal Stem Cells	13.9
97487_Patient-09ut_uterus	6.2	94730_Donor 3 AM - A_adipose	17.1
97488_Patient-09pl_placenta	3.3	94731_Donor 3 AM - B_adipose	11.7
97492_Patient-10ut_uterus	1.8	94732_Donor 3 AM - C_adipose	10.7
97493_Patient-10pl_placenta	14.0	94733_Donor 3 AD - A_adipose	85.9
97495_Patient-11go_adipose	14.4	94734_Donor 3 AD - B_adipose	19.2
97496_Patient-11sk_skeletal muscle	5.9	94735_Donor 3 AD - C_adipose	36.1
97497_Patient-11ut_uterus	1.8	77138_Liver_HepG2untreated	97.3
97498_Patient-11pl_placenta	6.0	73556_Heart_Cardiac stromal cells (primary)	9.3
97500_Patient-12go_adipose	21.9	81735_Small Intestine	78.5
97501_Patient-12sk_skeletal muscle	100.0	72409_Kidney_Proximal Convoluted Tubule	20.4
97502_Patient-12ut_uterus	3.3	82685_Small intestine_Duodenum	41.2
97503_Patient-12pl_placenta	3.2	90650_Adrenal_Adrenocortical adenoma	17.4
94721_Donor 2 U - A_Mesenchymal Stem Cells	2.5	72410_Kidney_HRCE	52.5
94722_Donor 2 U - B_Mesenchymal	2.4	72411_Kidney_HRE	25.7

Stem Cells			
94723_Donor 2 U - C_Mesenchymal Stem Cells	3.4	73139_Uterus_Uterine smooth muscle cells	14.4

**General\_screening\_panel\_v1.4 Summary:** Ag4830. Highest expression of this gene is seen in a renal cancer cell line (CT=26.2). This gene is widely expressed in this panel, with high to moderate expression seen in brain, colon, gastric, lung, breast, ovarian, and melanoma cancer cell lines. This expression profile suggests a role for this gene product in cell survival and proliferation. Modulation of this gene product may be useful in the treatment of cancer.

Among tissues with metabolic function, this gene is expressed at high to moderate levels in pituitary, adipose, adrenal gland, pancreas, thyroid, and adult and fetal skeletal muscle, heart, and liver. This widespread expression among these tissues suggests that this gene product may play a role in normal neuroendocrine and metabolic function and that disregulated expression of this gene may contribute to neuroendocrine disorders or metabolic diseases, such as obesity and diabetes. This gene encodes acetyl coA synthase. Inhibiting the production of acetyl CoA from one pathway may increase the utilization (energy generation) of acetyl CoA produced from other pathways. Decreased acetyl CoA will be available for lipid synthesis. Therefore, an inhibitor of ACS may facilitate weight loss and prevent weight gain, and be useful in the treatment of obesity.

In addition, this gene is expressed at much higher levels in fetal liver tissue (CT=27) when compared to expression in the adult counterpart (CT=30). Thus, expression of this gene may be used to differentiate between the fetal and adult source of this tissue.

This gene is also expressed at moderate levels in the CNS, including the hippocampus, thalamus, substantia nigra, amygdala, cerebellum and cerebral cortex. Therefore, therapeutic modulation of the expression or function of this gene may be useful in the treatment of neurologic disorders, such as Alzheimer's disease, Parkinson's disease, schizophrenia, multiple sclerosis, stroke and epilepsy.

**Panel 5 Islet Summary:** Ag4830. Highest expression of this gene is seen in diabetic skeletal muscle (CT=29) (patient 12). This gene is also expressed in other metabolic

tissues, including adipose and placenta. Please see Panel 1.4 for discussion of utility of this gene in metabolic disease.

#### H. CG124907-01: ornithine decarboxylase.

Expression of gene CG124907-01 was assessed using the primer-probe set Ag4751,  
5 described in Table HA. Results of the RTQ-PCR runs are shown in Tables HB and HC.

Table HA. Probe Name Ag4751

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-ctggatctgaggatgtgaaact-3'	22	894	486
Probe	TET-5'-cgtaatcaaccagcggttgacaaat-3'-TAMRA	26	937	487
Reverse	5'-actccagagtctgacggaaagt-3'	22	963	488

Table HB. General\_screening\_panel\_v1.4

Tissue Name	Rel. Exp.(%) Ag4751, Run 219997032	Tissue Name	Rel. Exp.(%) Ag4751, Run 219997032
Adipose	5.2	Renal ca. TK-10	17.7
Melanoma* Hs688(A).T	6.7	Bladder	8.8
Melanoma* Hs688(B).T	8.8	Gastric ca. (liver met.) NCI-N87	18.7
Melanoma* M14	5.4	Gastric ca. KATO III	85.3
Melanoma* LOXIMVI	22.1	Colon ca. SW-948	11.7
Melanoma* SK- MEL-5	32.5	Colon ca. SW480	49.7
Squamous cell carcinoma SCC-4	10.1	Colon ca.* (SW480 met) SW620	37.4
Testis Pool	6.9	Colon ca. HT29	17.8
Prostate ca.* (bone met) PC-3	100.0	Colon ca. HCT-116	68.3
Prostate Pool	2.8	Colon ca. CaCo-2	27.2
Placenta	0.3	Colon cancer tissue	10.3
Uterus Pool	1.8	Colon ca. SW1116	4.7
Ovarian ca. OVCAR-3	24.7	Colon ca. Colo-205	6.4
Ovarian ca. SK- OV-3	10.0	Colon ca. SW-48	6.6

Ovarian ca. OVCAR-4	7.3	Colon Pool	3.7
Ovarian ca. OVCAR-5	9.2	Small Intestine Pool	2.2
Ovarian ca. IGROV-1	18.8	Stomach Pool	2.2
Ovarian ca. OVCAR-8	6.5	Bone Marrow Pool	1.4
Ovary	1.5	Fetal Heart	2.0
Breast ca. MCF-7	10.7	Heart Pool	2.1
Breast ca. MDA-MB-231	17.3	Lymph Node Pool	2.8
Breast ca. BT 549	13.4	Fetal Skeletal Muscle	1.8
Breast ca. T47D	17.9	Skeletal Muscle Pool	6.3
Breast ca. MDA-N	2.5	Spleen Pool	1.4
Breast Pool	4.1	Thymus Pool	2.7
Trachea	2.7	CNS cancer (glio/astro) U87-MG	24.0
Lung	1.0	CNS cancer (glio/astro) U-118-MG	66.4
Fetal Lung	6.0	CNS cancer (neuro;met) SK-N-AS	6.0
Lung ca. NCI-N417	14.7	CNS cancer (astro) SF-539	7.9
Lung ca. LX-1	22.5	CNS cancer (astro) SNB-75	8.5
Lung ca. NCI-H146	14.3	CNS cancer (glio) SNB-19	15.9
Lung ca. SHP-77	54.0	CNS cancer (glio) SF-295	21.5
Lung ca. A549	13.3	Brain (Amygdala) Pool	1.4
Lung ca. NCI-H526	27.9	Brain (cerebellum)	2.3
Lung ca. NCI-H23	29.1	Brain (fetal)	9.5
Lung ca. NCI-H460	29.1	Brain (Hippocampus) Pool	1.8
Lung ca. HOP-62	4.9	Cerebral Cortex Pool	1.9
Lung ca. NCI-H522	31.2	Brain (Substantia nigra) Pool	1.4
Liver	0.6	Brain (Thalamus) Pool	1.8
Fetal Liver	8.8	Brain (whole)	2.6
Liver ca. HepG2	17.3	Spinal Cord Pool	1.8
Kidney Pool	4.4	Adrenal Gland	1.9
Fetal Kidney	16.6	Pituitary gland Pool	1.0

Renal ca. 786-0	5.8	Salivary Gland	1.0
Renal ca. A498	1.7	Thyroid (female)	7.0
Renal ca. ACHN	5.9	Pancreatic ca. CAPAN2	4.2
Renal ca. UO-31	10.2	Pancreas Pool	4.2

Table HC. Panel 5D

Tissue Name	Rel. Exp.(%) Ag4751, Run 204263059	Tissue Name	Rel. Exp.(%) Ag4751, Run 204263059
97457_Patient-02go_adipose	9.2	94709_Donor 2 AM - A_adipose	29.9
97476_Patient-07sk_skeletal muscle	7.3	94710_Donor 2 AM - B_adipose	22.1
97477_Patient-07ut_uterus	11.3	94711_Donor 2 AM - C_adipose	17.3
97478_Patient-07pl_placenta	1.5	94712_Donor 2 AD - A_adipose	30.8
97481_Patient-08sk_skeletal muscle	8.1	94713_Donor 2 AD - B_adipose	41.2
97482_Patient-08ut_uterus	10.9	94714_Donor 2 AD - C_adipose	39.2
97483_Patient-08pl_placenta	0.2	94742_Donor 3 U - A_Mesenchymal Stem Cells	9.0
97486_Patient-09sk_skeletal muscle	3.2	94743_Donor 3 U - B_Mesenchymal Stem Cells	28.1
97487_Patient-09ut_uterus	9.9	94730_Donor 3 AM - A_adipose	32.1
97488_Patient-09pl_placenta	3.0	94731_Donor 3 AM - B_adipose	17.6
97492_Patient-10ut_uterus	12.4	94732_Donor 3 AM - C_adipose	17.0
97493_Patient-10pl_placenta	3.9	94733_Donor 3 AD - A_adipose	45.4
97495_Patient-11go_adipose	4.0	94734_Donor 3 AD - B_adipose	23.8
97496_Patient-11sk_skeletal muscle	8.0	94735_Donor 3 AD - C_adipose	38.4
97497_Patient-11ut_uterus	25.2	77138_Liver_HepG2untreated	100.0
97498_Patient-11pl_placenta	1.2	73556_Heart_Cardiac stromal cells (primary)	11.7
97500_Patient-12go_adipose	12.6	81735_Small Intestine	10.0

97501_Patient-12sk_skeletal muscle	30.6	72409_Kidney_Proximal Convoluted Tubule	11.8
97502_Patient-12ut_uterus	21.8	82685_Small intestine_Duodenum	6.5
97503_Patient-12pl_placenta	1.5	90650_Adrenal_Adrenocortical adenoma	1.5
94721_Donor 2 U - A_Mesenchymal Stem Cells	29.9	72410_Kidney_HRCE	42.6
94722_Donor 2 U - B_Mesenchymal Stem Cells	21.3	72411_Kidney_HRE	41.5
94723_Donor 2 U - C_Mesenchymal Stem Cells	23.8	73139_Uterus_Uterine smooth muscle cells	19.2

**General\_screening\_panel\_v1.4 Summary:** Ag4751 Highest expression of this gene is detected in prostate cancer PC3 cell line (CT=23.5). High expression of this gene is also seen in cluster of cancer cell lines derived from pancreatic, gastric, colon, lung, liver, renal, breast, ovarian, prostate, squamous cell carcinoma, melanoma and brain cancers. Thus, expression of this gene could be used as a marker to detect the presence of these cancers. Furthermore, therapeutic modulation of the expression or function of this gene may be effective in the treatment of pancreatic, gastric, colon, lung, liver, renal, breast, ovarian, prostate, squamous cell carcinoma, melanoma and brain cancers.

Among tissues with metabolic or endocrine function, this gene is expressed at high to moderate levels in pancreas, adipose, adrenal gland, thyroid, pituitary gland, skeletal muscle, heart, liver and the gastrointestinal tract. Therefore, therapeutic modulation of the activity of this gene may prove useful in the treatment of endocrine/metabolically related diseases, such as obesity and diabetes.

This gene codes for ornithine Decarboxylase 1 (ODC). ODC is one of the key enzymes in polyamine biosynthesis. Preventing the accumulation of polyamines and their antilipolytic effects by inhibition of ODC at an earlier stage of obesity may inhibit progression of the obesity. In multiple GeneCalling studies at Curagen, enzyme spermidine/spermine acetyl transferase is found to be dysregulated in various disease models. This enzyme is one of the rate-limiting enzymes in the production of polyamines, spermidine and spermine. Previously, it was shown that oxidation of polyamines leads to generation of hydrogen

peroxide, which has been shown to have antilipolytic effects on adipose and may be involved in the progression of obesity.

In addition, this gene is expressed at high levels in all regions of the central nervous system examined, including amygdala, hippocampus, substantia nigra, thalamus, cerebellum, cerebral cortex, and spinal cord. Therefore, therapeutic modulation of this gene product may be useful in the treatment of central nervous system disorders such as Alzheimer's disease, Parkinson's disease, epilepsy, multiple sclerosis, schizophrenia and depression.

Interestingly, this gene is expressed at much higher levels in fetal (CT=27) when compared to adult liver (CT=31). This observation suggests that expression of this gene can be used to distinguish fetal from adult liver. In addition, the relative overexpression of this gene in fetal tissue suggests that the protein product may enhance liver growth or development in the fetus and thus may also act in a regenerative capacity in the adult. Therefore, therapeutic modulation of the protein encoded by this gene could be useful in treatment of liver related diseases.

#### References:

- 1: Taylor JL, Turo KA, McCann PP, Grossberg SE. Inhibition of the differentiation of 3T3-L1 cells by interferon-beta and difluoromethyl ornithine. *J. Biol. Regul. Homeost. Agents* 1988 Jan-Mar;2(1):19-24. PMID: 3140600.
- 2: Brown AP, Morrissey RL, Crowell JA, Levine BS. Difluoromethylornithine in combination with tamoxifen in female rats: 13-week oral toxicity study. *Cancer Chemother Pharmacol* 1999;44(6):475-83. PMID: 10550568.
- 3: Olefsky JM. Comparison of the effects of insulin and insulin-like agents on different aspects of adipocyte metabolism. *Horm. Metab. Res.* 1979 Mar;11(3):209-13. PMID: 447201.
- 4: Richelsen B, Pedersen SB, Hougaard DM. Characterization of antilipolytic action of polyamines in isolated rat adipocytes. *Biochem. J.* 1989 Jul 15;261(2):661-5. PMID: 2476118.
- 5: Livingston JN, Gurny PA, Lockwood DH. Insulin-like effects of polyamines in fat cells. Mediation by H<sub>2</sub>O<sub>2</sub> formation. *J. Biol. Chem.* 1977 Jan 25;252(2):560-2. PMID: 833144.

**Panel 5D Summary:** Ag4751 Highest expression of this gene is detected in liver cancer HepG2 cell line (CT=29.5). This gene shows a wide spread expression in this panel, which correlates with the expression in panel 1.4. Moderate expression of this gene is detected in adipose, skeletal muscle, uterus, placenta, heart smooth muscle, small intestine and kidney.

- 5 Therefore, therapeutic modulation of this gene may be useful in the treatment of obesity and diabetes including type II diabetes.

#### **I. CG128347-02: kinesin-like.**

Expression of gene CG128347-02 was assessed using the primer-probe set Ag5691, described in Table IA. Results of the RTQ-PCR runs are shown in Table IB.

10 **Table IA.** Probe Name Ag5691

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-gaattagacctctgctttgcaa-3'	22	164	489
Probe	TET-5'-cacacaaacttgatgattatgaagagcttc-3'-TAMRA	30	187	490
Reverse	5'-gctggctgtttggaataactct-3'	22	217	491

**Table IB.** Panel 4.1D

Tissue Name	Rel. Exp.(%) Ag5691, Run 246504797	Tissue Name	Rel. Exp.(%) Ag5691, Run 246504797
Secondary Th1 act	9.8	HUVEC IL-1beta	8.2
Secondary Th2 act	23.0	HUVEC IFN gamma	9.7
Secondary Tr1 act	5.0	HUVEC TNF alpha + IFN gamma	0.0
Secondary Th1 rest	0.0	HUVEC TNF alpha + IL4	2.8
Secondary Th2 rest	0.0	HUVEC IL-11	6.4
Secondary Tr1 rest	0.0	Lung Microvascular EC none	20.7
Primary Th1 act	0.0	Lung Microvascular EC TNFalpha + IL-1beta	3.0
Primary Th2 act	11.9	Microvascular Dermal EC none	1.7
Primary Tr1 act	10.2	Microvascular Dermal EC TNFalpha + IL-1beta	3.4
Primary Th1 rest	0.0	Bronchial epithelium TNFalpha + IL1beta	11.0
Primary Th2 rest	2.3	Small airway epithelium	6.1

		none	
Primary Tr1 rest	0.0	Small airway epithelium TNFalpha + IL-1beta	9.6
CD45RA CD4 lymphocyte act	8.4	Coronary artery SMC rest	3.6
CD45RO CD4 lymphocyte act	13.8	Coronary artery SMC TNFalpha + IL-1beta	7.9
CD8 lymphocyte act	0.0	Astrocytes rest	1.1
Secondary CD8 lymphocyte rest	9.2	Astrocytes TNFalpha + IL-1beta	0.0
Secondary CD8 lymphocyte act	0.6	KU-812 (Basophil) rest	13.7
CD4 lymphocyte none	0.9	KU-812 (Basophil) PMA/ionomycin	11.5
2ry Th1/Th2/Tr1_anti- CD95 CH11	0.0	CCD1106 (Keratinocytes) none	25.0
LAK cells rest	5.9	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	12.6
LAK cells IL-2	3.3	Liver cirrhosis	9.6
LAK cells IL-2+IL-12	1.4	NCI-H292 none	15.5
LAK cells IL-2+IFN gamma	2.5	NCI-H292 IL-4	17.8
LAK cells IL-2+ IL-18	1.5	NCI-H292 IL-9	39.0
LAK cells PMA/ionomycin	3.4	NCI-H292 IL-13	28.3
NK Cells IL-2 rest	1.5	NCI-H292 IFN gamma	2.8
Two Way MLR 3 day	4.8	HPAEC none	3.8
Two Way MLR 5 day	0.0	HPAEC TNF alpha + IL- 1 beta	18.7
Two Way MLR 7 day	1.6	Lung fibroblast none	7.6
PBMC rest	0.3	Lung fibroblast TNF alpha + IL-1 beta	9.0
PBMC PWM	0.8	Lung fibroblast IL-4	12.5
PBMC PHA-L	2.2	Lung fibroblast IL-9	6.8
Ramos (B cell) none	2.2	Lung fibroblast IL-13	1.6
Ramos (B cell) ionomycin	18.6	Lung fibroblast IFN gamma	5.9
B lymphocytes PWM	10.5	Dermal fibroblast CCD1070 rest	10.1
B lymphocytes CD40L and IL-4	15.1	Dermal fibroblast CCD1070 TNF alpha	0.0
EOL-1 dbcAMP	2.8	Dermal fibroblast CCD1070 IL-1 beta	5.4

EOL-1 dbcAMP PMA/ionomycin	0.0	Dermal fibroblast IFN gamma	3.3
Dendritic cells none	3.2	Dermal fibroblast IL-4	14.2
Dendritic cells LPS	1.1	Dermal Fibroblasts rest	6.6
Dendritic cells anti- CD40	0.0	Neutrophils TNFa+LPS	6.7
Monocytes rest	0.5	Neutrophils rest	100.0
Monocytes LPS	18.6	Colon	1.1
Macrophages rest	3.3	Lung	0.4
Macrophages LPS	0.0	Thymus	10.0
HUVEC none	5.2	Kidney	28.3
HUVEC starved	2.4		

**CNS\_neurodegeneration\_v1.0 Summary:** Ag5691 Results from one experiment with this gene are not included. The amp plot indicates that there were experimental difficulties with this run (Data not shown).

5. **General\_screening\_panel\_v1.5 Summary:** Ag5691 Results from one experiment with this gene are not included. The amp plot indicates that there were experimental difficulties with this run (Data not shown).

10. **Panel 4.1D Summary:** AG5691 Highest expression of this gene is seen in resting neutrophils (CT=31.3). This expression is reduced to background level (CT=35.2) in neutrophils activated by TNF-alpha+LPS. This expression profile suggests that the protein encoded by this gene is produced by resting neutrophils but not by activated neutrophils. Therefore, the gene product may reduce activation of these inflammatory cells and modulation of its expression or activity may reduce or eliminate the symptoms in patients with Crohn's disease, ulcerative colitis, multiple sclerosis, chronic obstructive pulmonary disease, asthma, emphysema, rheumatoid arthritis, lupus erythematosus, or psoriasis.
15. In addition, antagonists of this gene product may be effective in increasing the immune response in patients with AIDS or other immunodeficiencies.

#### **J. CG135823-01 and CG135823-02: TAT.**

20. Expression of gene CG135823-01 and CG135823-02 was assessed using the primer-probe sets Ag3173 and Ag4906, described in Tables JA and JB. Results of the RTQ-PCR runs are shown in Tables JC and JD. Please note that probe-primer set Ag4906 is specific for CG135823-01 variant.

Table JA. Probe Name Ag3173

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-ctctggctgagtctatgggaat-3'	22	617	492
Probe	TET-5'-tgaggtcaaactctacaatttgttgcca-3'- TAMRA	28	639	493
Reverse	5'-tcaggtcaatttcccaagattt-3'	22	670	494

Table JB. Probe Name Ag4906

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-ctcaggatgagggaaaagaaaa-3'	22	1796	495
Probe	TET-5'-ccccaaccatttctcagactcta-3'- TAMRA	24	1837	496
Reverse	5'-tggagagagcgtgttctttct-3'	21	1861	497

Table JC. General\_screening\_panel\_v1.5

Tissue Name	Rel. Exp.(%) Ag4906, Run 228783186	Tissue Name	Rel. Exp.(%) Ag4906, Run 228783186
Adipose	0.1	Renal ca. TK-10	3.4
Melanoma* Hs688(A).T	0.1	Bladder	0.3
Melanoma* Hs688(B).T	0.0	Gastric ca. (liver met.) NCI-N87	0.1
Melanoma* M14	0.0	Gastric ca. KATO III	0.0
Melanoma* LOXIMVI	0.0	Colon ca. SW-948	0.0
Melanoma* SK- MEL-5	0.0	Colon ca. SW480	0.1
Squamous cell carcinoma SCC-4	0.0	Colon ca.* (SW480 met) SW620	0.0
Testis Pool	0.5	Colon ca. HT29	0.0
Prostate ca.* (bone met) PC-3	0.0	Colon ca. HCT-116	0.1
Prostate Pool	0.0	Colon ca. CaCo-2	0.1
Placenta	0.0	Colon cancer tissue	0.0
Uterus Pool	0.1	Colon ca. SW1116	0.0
Ovarian ca. OVCAR-3	0.1	Colon ca. Colo-205	0.0
Ovarian ca. SK- OV-3	0.0	Colon ca. SW-48	0.0

Ovarian ca. OVCAR-4	0.0	Colon Pool	0.1
Ovarian ca. OVCAR-5	0.3	Small Intestine Pool	0.0
Ovarian ca. IGROV-1	0.0	Stomach Pool	0.2
Ovarian ca. OVCAR-8	0.0	Bone Marrow Pool	0.0
Ovary	0.1	Fetal Heart	0.0
Breast ca. MCF-7	0.1	Heart Pool	0.0
Breast ca. MDA-MB-231	0.0	Lymph Node Pool	0.1
Breast ca. BT 549	0.0	Fetal Skeletal Muscle	0.0
Breast ca. T47D	0.1	Skeletal Muscle Pool	0.0
Breast ca. MDA-N	0.0	Spleen Pool	0.0
Breast Pool	0.0	Thymus Pool	0.1
Trachea	0.1	CNS cancer (glio/astro) U87-MG	0.0
Lung	0.0	CNS cancer (glio/astro) U-118-MG	0.0
Fetal Lung	0.1	CNS cancer (neuro;met) SK-N-AS	0.0
Lung ca. NCI-N417	0.0	CNS cancer (astro) SF-539	0.0
Lung ca. LX-1	0.1	CNS cancer (astro) SNB-75	0.0
Lung ca. NCI-H146	0.0	CNS cancer (glio) SNB-19	0.0
Lung ca. SHP-77	0.1	CNS cancer (glio) SF-295	0.0
Lung ca. A549	0.5	Brain (Amygdala) Pool	0.0
Lung ca. NCI-H526	0.0	Brain (cerebellum)	0.0
Lung ca. NCI-H23	0.1	Brain (fetal)	0.0
Lung ca. NCI-H460	0.9	Brain (Hippocampus) Pool	0.1
Lung ca. HOP-62	0.0	Cerebral Cortex Pool	0.0
Lung ca. NCI-H522	0.1	Brain (Substantia nigra) Pool	0.0
Liver	100.0	Brain (Thalamus) Pool	0.0
Fetal Liver	8.2	Brain (whole)	1.0
Liver ca. HepG2	7.6	Spinal Cord Pool	0.0
Kidney Pool	0.0	Adrenal Gland	0.3
Fetal Kidney	0.1	Pituitary gland Pool	0.0

Renal ca. 786-0	0.0	Salivary Gland	0.0
Renal ca. A498	0.0	Thyroid (female)	0.0
Renal ca. ACHN	0.0	Pancreatic ca. CAPAN2	0.0
Renal ca. UO-31	0.0	Pancreas Pool	0.1

Table JD. Panel 5 Islet

Tissue Name	Rel. Exp.(%) Ag4906, Run 223846056	Tissue Name	Rel. Exp.(%) Ag4906, Run 223846056
97457_Patient-02go_adipose	0.0	94709_Donor 2 AM - A_adipose	0.2
97476_Patient-07sk_skeletal muscle	0.0	94710_Donor 2 AM - B_adipose	0.0
97477_Patient-07ut_uterus	0.0	94711_Donor 2 AM - C_adipose	0.0
97478_Patient-07pl_placenta	0.0	94712_Donor 2 AD - A_adipose	0.4
99167_Bayer Patient 1	0.0	94713_Donor 2 AD - B_adipose	0.6
97482_Patient-08ut_uterus	0.0	94714_Donor 2 AD - C_adipose	0.0
97483_Patient-08pl_placenta	0.0	94742_Donor 3 U - A Mesenchymal Stem Cells	0.0
97486_Patient-09sk_skeletal muscle	0.0	94743_Donor 3 U - B Mesenchymal Stem Cells	0.0
97487_Patient-09ut_uterus	0.0	94730_Donor 3 AM - A_adipose	0.6
97488_Patient-09pl_placenta	0.0	94731_Donor 3 AM - B_adipose	0.0
97492_Patient-10ut_uterus	0.6	94732_Donor 3 AM - C_adipose	0.0
97493_Patient-10pl_placenta	0.0	94733_Donor 3 AD - A_adipose	0.0
97495_Patient-11go_adipose	0.0	94734_Donor 3 AD - B_adipose	0.0
97496_Patient-11sk_skeletal muscle	0.0	94735_Donor 3 AD - C_adipose	0.0
97497_Patient-11ut_uterus	0.0	77138_Liver_HepG2untreated	100.0
97498_Patient-11pl_placenta	0.0	73556_Heart_Cardiac stromal cells (primary)	0.0
97500_Patient-12go_adipose	0.0	81735_Small Intestine	1.0

97501_Patient-12sk_skeletal muscle	0.0	72409_Kidney_Proximal Convoluted Tubule	0.0
97502_Patient-12ut_uterus	0.6	82685_Small intestine_Duodenum	0.7
97503_Patient-12pl_placenta	0.0	90650_Adrenal_Adrenocortical adenoma	3.1
94721_Donor 2 U - A_Mesenchymal Stem Cells	0.0	72410_Kidney_HRCE	0.0
94722_Donor 2 U - B_Mesenchymal Stem Cells	0.0	72411_Kidney_HRE	0.0
94723_Donor 2 U - C_Mesenchymal Stem Cells	0.0	73139_Uterus_Uterine smooth muscle cells	0.0

**General\_screening\_panel\_v1.5 Summary:** Ag4906 This gene seems to be almost exclusively expressed in liver (CT=24.6). A lower level of expression has been detected in fetal liver (CT=28) and brain. Thus, expression of this gene could be used to differentiate between liver and fetal liver tissues. In addition, the relative overexpression of this gene in fetal liver suggests that the protein product may enhance liver growth or development in the fetus and thus may also act in a regenerative capacity in the adult. Therefore, therapeutic modulation of the protein encoded by this gene could be useful in treatment of liver and metabolic related diseases, including obesity and diabetes.

**Panel 5 Islet Summary:** Ag4906 This gene is expressed in hepatocyte-derived HepG2 cell line (CT=29.8), which is in accordance with the liver expression seen in panel 1.5.

#### K. CG140122-01: Polyamine Oxidase.

Expression of gene CG140122-01 was assessed using the primer-probe sets Ag4986 and Ag5105, described in Tables KA and KB. Results of the RTQ-PCR runs are shown in Tables KC and KD.

Table KA. Probe Name Ag4986

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-gtgcagagtgtgaaacttgg-3'	21	259	498

Probe	TET-5'-catggctcccatgggaaccctat-3'- TAMRA	23	313	499
Reverse	5'-cgttggtctctgctagatgata-3'	22	337	500

Table KB. Probe Name Ag5105

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-gaccgtgtcgctaggt-3'	16	1059	501
Probe	TET-5'-cagtaaccagtttcttccggcca-3'- TAMRA	24	1087	502
Reverse	5'-accttctctgtgggcag-3'	17	1114	503

Table KC. CNS\_neurodegeneration\_v1.0

Tissue Name	Rel. Exp.(%) Ag5105, Run 249286379	Tissue Name	Rel. Exp.(%) Ag5105, Run 249286379
AD 1 Hippo	27.5	Control (Path) 3 Temporal Ctx	12.2
AD 2 Hippo	50.7	Control (Path) 4 Temporal Ctx	20.6
AD 3 Hippo	18.9	AD 1 Occipital Ctx	23.7
AD 4 Hippo	17.1	AD 2 Occipital Ctx (Missing)	0.0
AD 5 hippo	63.7	AD 3 Occipital Ctx	18.8
AD 6 Hippo	100.0	AD 4 Occipital Ctx	18.8
Control 2 Hippo	35.4	AD 5 Occipital Ctx	13.8
Control 4 Hippo	24.3	AD 6 Occipital Ctx	28.3
Control (Path) 3 Hippo	10.6	Control 1 Occipital Ctx	12.0
AD 1 Temporal Ctx	36.3	Control 2 Occipital Ctx	39.0
AD 2 Temporal Ctx	21.2	Control 3 Occipital Ctx	23.0
AD 3 Temporal Ctx	20.2	Control 4 Occipital Ctx	18.6
AD 4 Temporal Ctx	20.9	Control (Path) 1. Occipital Ctx	39.2
AD 5 Inf Temporal Ctx	50.0	Control (Path) 2 Occipital Ctx	8.6

AD 5 SupTemporal Ctx	64.6	Control (Path) 3 Occipital Ctx	10.3
AD 6 Inf Temporal Ctx	58.6	Control (Path) 4 Occipital Ctx	9.8
AD 6 Sup Temporal Ctx	39.5	Control 1 Parietal Ctx	17.2
Control 1 Temporal Ctx	14.9	Control 2 Parietal Ctx	69.3
Control 2 Temporal Ctx	32.3	Control 3 Parietal Ctx	17.9
Control 3 Temporal Ctx	19.3	Control (Path) 1 Parietal Ctx	42.0
Control 4 Temporal Ctx	21.8	Control (Path) 2 Parietal Ctx	20.0
Control (Path) 1 Temporal Ctx	21.0	Control (Path) 3 Parietal Ctx	11.0
Control (Path) 2 Temporal Ctx	19.8	Control (Path) 4 Parietal Ctx	11.2

Table KD. General\_screening\_panel\_v1.5

Tissue Name	Rel. Exp.(%) Ag5105, Run 228969349	Rel. Exp.(%) Ag5105, Run 229514472	Tissue Name	Rel. Exp.(%) Ag5105, Run 228969349	Rel. Exp.(%) Ag5105, Run 229514472
Adipose	1.9	1.4	Renal ca. TK-10	26.8	29.7
Melanoma* Hs688(A).T	2.8	2.6	Bladder	2.9	3.6
Melanoma* Hs688(B).T	2.7	2.4	Gastric ca. (liver met.) NCI-N87	13.0	12.8
Melanoma* M14	2.2	2.1	Gastric ca. KATO III	14.4	17.2
Melanoma* LOXIMVI	9.9	10.7	Colon ca. SW-948	4.2	3.7
Melanoma* SK-MEL-5	5.9	5.8	Colon ca. SW480	11.3	10.3
Squamous cell carcinoma SCC-4	4.0	2.8	Colon ca.* (SW480 met) SW620	22.7	24.1
Testis Pool	2.0	1.8	Colon ca. HT29	5.6	5.8
Prostate ca.* (bone met) PC-3	33.9	42.9	Colon ca. HCT-116	9.5	11.9
Prostate Pool	1.8	1.8	Colon ca. CaCo-2	15.5	18.3

Placenta	0.5	0.5	Colon cancer tissue	8.8	11.8
Uterus Pool	1.3	1.6	Colon ca. SW1116	1.9	1.0
Ovarian ca. OVCAR-3	1.8	2.1	Colon ca. Colo-205	7.2	8.5
Ovarian ca. SK-OV-3	7.2	9.9	Colon ca. SW-48	6.3	5.5
Ovarian ca. OVCAR-4	1.2	2.2	Colon Pool	1.7	1.7
Ovarian ca. OVCAR-5	17.0	21.3	Small Intestine Pool	2.5	2.7
Ovarian ca. IGROV-1	13.2	16.7	Stomach Pool	2.0	2.2
Ovarian ca. OVCAR-8	7.1	5.9	Bone Marrow Pool	1.6	1.6
Ovary	1.0	1.4	Fetal Heart	0.9	0.7
Breast ca. MCF-7	1.5	1.6	Heart Pool	0.3	0.8
Breast ca. MDA-MB-231	5.1	5.4	Lymph Node Pool	3.2	2.6
Breast ca. BT 549	14.5	13.3	Fetal Skeletal Muscle	0.6	0.4
Breast ca. T47D	0.1	0.0	Skeletal Muscle Pool	0.6	1.1
Breast ca. MDA-N	2.1	2.7	Spleen Pool	0.9	1.1
Breast Pool	2.6	2.1	Thymus Pool	2.0	2.3
Trachea	2.6	2.3	CNS cancer (glio/astro) U87-MG	8.2	9.7
Lung	0.5	0.5	CNS cancer (glio/astro) U-118-MG	12.2	13.6
Fetal Lung	2.2	2.9	CNS cancer (neuro;met) SK-N-AS	1.7	1.7
Lung ca. NCI-N417	0.1	0.1	CNS cancer (astro) SF-539	1.5	1.8
Lung ca. LX-1	18.2	20.0	CNS cancer (astro) SNB-75	8.3	18.4
Lung ca. NCI-H146	0.0	0.0	CNS cancer (glio) SNB-19	17.8	19.6
Lung ca.	0.7	0.6	CNS cancer	15.0	15.9

SHP-77			(glio) SF-295		
Lung ca. A549	33.4	36.9	Brain (Amygdala) Pool	5.1	5.4
Lung ca. NCI-H526	2.7	3.0	Brain (cerebellum)	7.5	10.2
Lung ca. NCI-H23	3.1	3.2	Brain (fetal)	4.2	5.6
Lung ca. NCI-H460	100.0	100.0	Brain (Hippocampus) Pool	8.3	6.8
Lung ca. HOP-62	6.0	6.0	Cerebral Cortex Pool	6.5	5.3
Lung ca. NCI-H522	3.8	4.9	Brain (Substantia nigra) Pool	8.5	7.0
Liver	0.2	0.2	Brain (Thalamus) Pool	7.4	8.4
Fetal Liver	3.3	3.7	Brain (whole)	6.3	6.3
Liver ca. HepG2	7.2	7.0	Spinal Cord Pool	11.4	12.6
Kidney Pool	2.5	2.8	Adrenal Gland	0.9	1.0
Fetal Kidney	2.0	2.0	Pituitary gland Pool	0.3	0.2
Renal ca. 786-0	13.4	13.7	Salivary Gland	1.6	1.7
Renal ca. A498	2.3	2.2	Thyroid (female)	0.7	1.1
Renal ca. ACHN	4.0	5.1	Pancreatic ca. CAPAN2	13.0	14.7
Renal ca. UO-31	5.7	6.2	Pancreas Pool	2.9	3.8

**CNS\_neurodegeneration\_v1.0 Summary:** Ag5105 This panel confirms the expression of this gene at low levels in the brain in an independent group of individuals. This gene is found to be upregulated in the temporal cortex of Alzheimer's disease patients. Therefore, therapeutic modulation of the expression or function of this gene may decrease neuronal death and be of use in the treatment of this disease.

**General\_screening\_panel\_v1.4 Summary:** Ag4986 Expression of this gene is low/undetectable in all samples on this panel (CTs>35). (Data not shown.)

**General\_screening\_panel\_v1.5 Summary:** Ag5105 Two experiments with the same probe and primer set produce results that are in excellent agreement. Highest expression of

this gene is seen in a breast cancer cell line (CTs=24-26). This gene is widely expressed in this panel, with high to moderate expression seen in brain, colon, gastric, lung, breast, ovarian, and melanoma cancer cell lines. This expression profile suggests a role for this gene product in cell survival and proliferation. Modulation of this gene product may be  
5 useful in the treatment of cancer.

Among tissues with metabolic function, this gene is expressed at moderate levels in pituitary, adipose, adrenal gland, pancreas, thyroid, and adult and fetal skeletal muscle, heart, and liver. This widespread expression among these tissues suggests that this gene product may play a role in normal neuroendocrine and metabolic function and that  
10 disregulated expression of this gene may contribute to neuroendocrine disorders or metabolic diseases, such as obesity and diabetes.

This gene is also expressed at moderate levels in the CNS, including the hippocampus, thalamus, substantia nigra, amygdala, cerebellum and cerebral cortex. Therefore, therapeutic modulation of the expression or function of this gene may be useful in the  
15 treatment of neurologic disorders, such as Alzheimer's disease, Parkinson's disease, schizophrenia, multiple sclerosis, stroke and epilepsy.

**Panel 5 Islet Summary:** Ag4986 Expression of this gene is low/undetectable in all samples on this panel (CTs>35). (Data not shown.)

**Panel 5D Summary:** Ag5105. Results from one experiment with this gene are not  
20 included. The amp plot indicates that there were experimental difficulties with this run.

**L. CG140316-01: Malic enzyme isoform1 (MB\_X77244 ).**

Expression of gene CG140316-01. was assessed using the primer-probe set Ag4998, described in Table LA. Results of the RTQ-PCR runs are shown in Tables LB and LC.

Table LA. Probe Name Ag4998

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-agtttgcccatgaacatgaa-3'	20	1058	504
Probe	TET-5'-gccattgttcaagaaataaaaccaactgc-3'-TAMRA	29	1096	505
Reverse	5'-ttgcagcaactcctatgagg-3'	20	1125	506

Table LB. General\_screening\_panel\_v1.4

Tissue Name	Rel. Exp.(%) Ag4998, Run 219998185	Tissue Name	Rel. Exp.(%) Ag4998, Run 219998185
Adipose	12.8	Renal ca. TK-10	7.6
Melanoma* Hs688(A).T	15.8	Bladder	3.9
Melanoma* Hs688(B).T	28.7	Gastric ca. (liver met.) NCI-N87	11.7
Melanoma* M14	8.7	Gastric ca. KATO III	36.3
Melanoma* LOXIMVI	9.9	Colon ca. SW-948	12.5
Melanoma* SK- MEL-5	22.2	Colon ca. SW480	26.1
Squamous cell carcinoma SCC-4	20.7	Colon ca.* (SW480 met) SW620	12.2
Testis Pool	7.2	Colon ca. HT29	21.3
Prostate ca.* (bone met) PC-3	100.0	Colon ca. HCT-116	59.0
Prostate Pool	2.8	Colon ca. CaCo-2	56.3
Placenta	0.2	Colon cancer tissue	7.9
Uterus Pool	0.9	Colon ca. SW1116	4.9
Ovarian ca. OVCAR-3	7.4	Colon ca. Colo-205	8.1
Ovarian ca. SK- OV-3	37.6	Colon ca. SW-48	4.5
Ovarian ca. OVCAR-4	10.7	Colon Pool	4.2
Ovarian ca. OVCAR-5	6.9	Small Intestine Pool	1.0
Ovarian ca. IGROV-1	4.0	Stomach Pool	1.9
Ovarian ca. OVCAR-8	6.0	Bone Marrow Pool	2.3
Ovary	6.4	Fetal Heart	2.3
Breast ca. MCF-7	12.6	Heart Pool	2.0
Breast ca. MDA- MB-231	16.2	Lymph Node Pool	3.0
Breast ca. BT 549	19.8	Fetal Skeletal Muscle	0.0
Breast ca. T47D	11.7	Skeletal Muscle Pool	8.8
Breast ca. MDA-N	0.0	Spleen Pool	3.0
Breast Pool	3.1	Thymus Pool	1.5
Trachea	5.6	CNS cancer	0.0

		(glio/astro) U87-MG	
Lung	1.3	CNS cancer (glio/astro) U-118-MG	10.7
Fetal Lung	5.4	CNS cancer (neuro;met) SK-N-AS	15.9
Lung ca. NCI-N417	0.8	CNS cancer (astro) SF-539	18.3
Lung ca. LX-1	8.3	CNS cancer (astro) SNB-75	0.1
Lung ca. NCI-H146	1.8	CNS cancer (glio) SNB-19	5.6
Lung ca. SHP-77	30.8	CNS cancer (glio) SF-295	0.0
Lung ca. A549	67.4	Brain (Amygdala) Pool	6.8
Lung ca. NCI-H526	1.7	Brain (cerebellum)	4.6
Lung ca. NCI-H23	6.2	Brain (fetal)	2.8
Lung ca. NCI-H460	55.9	Brain (Hippocampus) Pool	6.3
Lung ca. HOP-62	15.2	Cerebral Cortex Pool	9.3
Lung ca. NCI-H522	0.0	Brain (Substantia nigra) Pool	5.7
Liver	0.4	Brain (Thalamus) Pool	11.9
Fetal Liver	3.4	Brain (whole)	7.9
Liver ca. HepG2	0.0	Spinal Cord Pool	7.4
Kidney Pool	3.1	Adrenal Gland	26.4
Fetal Kidney	0.8	Pituitary gland Pool	3.6
Renal ca. 786-0	14.7	Salivary Gland	0.6
Renal ca. A498	14.2	Thyroid (female)	1.0
Renal ca. ACHN	20.3	Pancreatic ca. CAPAN2	9.3
Renal ca. UO-31	16.5	Pancreas Pool	2.7

Table LC. Panel 5D.

Tissue Name	Rel. Exp.(%) Ag4998, Run 220259861	Tissue Name	Rel. Exp.(%) Ag4998, Run 220259861
97457_Patient-02go_adipose	8.5	94709_Donor 2 AM - A_adipose	26.4
97476_Patient-07sk_skeletal muscle	5.2	94710_Donor 2 AM - B_adipose	11.7
97477_Patient-07ut_uterus	14.0	94711_Donor 2 AM - C_adipose	9.0
97478_Patient-	2.4	94712_Donor 2 AD - A_adipose	77.4

07pl_placenta			
97481_Patient-08sk_skeletal muscle	7.1	94713_Donor 2 AD - B_adipose	94.6
97482_Patient-08ut_uterus	9.7	94714_Donor 2 AD - C_adipose	100.0
97483_Patient-08pl_placenta	1.4	94742_Donor 3 U - A_Mesenchymal Stem Cells	6.7
97486_Patient-09sk_skeletal muscle	6.9	94743_Donor 3 U - B_Mesenchymal Stem Cells	12.4
97487_Patient-09ut_uterus	16.0	94730_Donor 3 AM - A_adipose	20.2
97488_Patient-09pl_placenta	1.2	94731_Donor 3 AM - B_adipose	16.6
97492_Patient-10ut_uterus	9.0	94732_Donor 3 AM - C_adipose	16.5
97493_Patient-10pl_placenta	3.5	94733_Donor 3 AD - A_adipose	92.7
97495_Patient-11go_adipose	5.9	94734_Donor 3 AD - B_adipose	55.1
97496_Patient-11sk_skeletal muscle	16.2	94735_Donor 3 AD - C_adipose	57.8
97497_Patient-11ut_uterus	23.0	77138_Liver_HepG2untreated	8.7
97498_Patient-11pl_placenta	0.0	73556_Heart_Cardiac stromal cells (primary)	9.0
97500_Patient-12go_adipose	28.9	81735_Small Intestine	5.0
97501_Patient-12sk_skeletal muscle	33.9	72409_Kidney_Proximal Convoluted Tubule	12.3
97502_Patient-12ut_uterus	15.4	82685_Small intestine_Duodenum	18.8
97503_Patient-12pl_placenta	0.3	90650_Adrenal_Adrenocortical adenoma	9.5
94721_Donor 2 U - A_Mesenchymal Stem Cells	10.2	72410_Kidney_HRCE	33.9
94722_Donor 2 U - B_Mesenchymal Stem Cells	36.1	72411_Kidney_HRE	25.3
94723_Donor 2 U - C_Mesenchymal Stem Cells	9.0	73139_Uterus_Uterine smooth muscle cells	19.2

**General\_screening\_panel\_v1.4 Summary:** Ag4998 Cytosolic malic enzyme is ubiquitously expressed including endocrine/metabolically-relevant tissues such as, adipose,

GI, liver, and skeletal muscle. These results indicate that this enzyme is critical for normal physiology. Furthermore, dysregulated expression of this gene may contribute to neuroendocrine disorders or metabolic diseases, such as obesity and diabetes.

Highest expression of this gene is seen in a prostate cancer cell line (CT=25.4). This gene is widely expressed in this panel, with high to moderate expression seen in brain, colon, gastric, lung, breast, ovarian, and melanoma cancer cell lines. This expression profile suggests a role for this gene product in cell survival and proliferation. Modulation of this gene product may be useful in the treatment of cancer.

This gene is also expressed at moderate levels in the CNS, including the hippocampus, thalamus, substantia nigra, amygdala, cerebellum and cerebral cortex. Therefore, therapeutic modulation of the expression or function of this gene may be useful in the treatment of neurologic disorders, such as Alzheimer's disease, Parkinson's disease, schizophrenia, multiple sclerosis, stroke and epilepsy.

**Panel 5D Summary:** Ag4998 Cytosolic malic enzyme has low to moderate expression in fully differentiated adipose, and adipose found in diabetic gestational diabetics.

#### M. CG142427-01: ATP citrate lyase.

Expression of gene CG142427-01 and CG142404-01 were assessed using the primer-probe set Ag6008, described in Table MA. Results of the RTQ-PCR runs are shown in Tables MB and MC.

Table MA. Probe Name Ag6008

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-agattacgtcaggcagcactt-3'	21	3113	507
Probe	TET-5'-cactcctctgctcgattatgcactgg-3'-TAMRA	26	3140	508
Reverse	5'-gcttcttcgaggtggtaatctt-3'	22	3174	509

Table MB. General\_screening\_panel\_v1.5

Tissue Name	Rel. Exp.(%) Ag6008. Run	Tissue Name	Rel. Exp.(%) Ag6008. Run
-------------	-----------------------------	-------------	-----------------------------

	228763479		228763479
Adipose	6.2	Renal ca. TK-10	64.2
Melanoma* Hs688(A).T	37.6	Bladder	12.4
Melanoma* Hs688(B).T	59.0	Gastric ca. (liver met.) NCI-N87	65.1
Melanoma* M14	55.9	Gastric ca. KATO III	59.5
Melanoma* LOXIMVI	59.0	Colon ca. SW-948	14.5
Melanoma* SK- MEL-5	41.8	Colon ca. SW480	62.4
Squamous cell carcinoma SCC-4	24.1	Colon ca.* (SW480 met) SW620	32.3
Testis Pool	6.0	Colon ca. HT29	27.4
Prostate ca.* (bone met) PC-3	32.8	Colon ca. HCT-116	45.7
Prostate Pool	13.0	Colon ca. CaCo-2	66.0
Placenta	6.1	Colon cancer tissue	8.3
Uterus Pool	6.6	Colon ca. SW1116	4.0
Ovarian ca. OVCAR-3	12.9	Colon ca. Colo-205	11.1
Ovarian ca. SK- OV-3	47.3	Colon ca. SW-48	14.9
Ovarian ca. OVCAR-4	17.2	Colon Pool	13.3
Ovarian ca. OVCAR-5	35.1	Small Intestine Pool	5.6
Ovarian ca. IGROV-1	22.2	Stomach Pool	4.0
Ovarian ca. OVCAR-8	8.2	Bone Marrow Pool	3.8
Ovary	8.0	Fetal Heart	3.5
Breast ca. MCF-7	23.7	Heart Pool	2.5
Breast ca. MDA- MB-231	46.7	Lymph Node Pool	8.4
Breast ca. BT 549	60.7	Fetal Skeletal Muscle	3.7
Breast ca. T47D	29.1	Skeletal Muscle Pool	3.4
Breast ca. MDA-N	12.9	Spleen Pool	5.3
Breast Pool	8.0	Thymus Pool	6.8
Trachea	9.3	CNS cancer (glio/astro) U87-MG	60.7
Lung	1.4	CNS cancer (glio/astro) U-118-MG	59.0
Fetal Lung	16.3	CNS cancer	60.7

		(neuro;met) SK-N-AS	
Lung ca. NCI-N417	30.1	CNS cancer (astro) SF-539	24.8
Lung ca. LX-1	28.1	CNS cancer (astro) SNB-75	32.5
Lung ca. NCI-H146	23.5	CNS cancer (glio) SNB-19	25.2
Lung ca. SHP-77	46.7	CNS cancer (glio) SF-295	76.8
Lung ca. A549	100.0	Brain (Amygdala) Pool	4.8
Lung ca. NCI-H526	10.0	Brain (cerebellum)	28.3
Lung ca. NCI-H23	23.5	Brain (fetal)	16.5
Lung ca. NCI-H460	25.5	Brain (Hippocampus) Pool	8.6
Lung ca. HOP-62	29.5	Cerebral Cortex Pool	10.5
Lung ca. NCI-H522	57.4	Brain (Substantia nigra) Pool	6.3
Liver	0.8	Brain (Thalamus) Pool	10.7
Fetal Liver	22.4	Brain (whole)	12.2
Liver ca. HepG2	23.0	Spinal Cord Pool	7.4
Kidney Pool	7.5	Adrenal Gland	13.2
Fetal Kidney	5.4	Pituitary gland Pool	1.9
Renal ca. 786-0	36.3	Salivary Gland	4.0
Renal ca. A498	33.0	Thyroid (female)	2.7
Renal ca. ACHN	80.7	Pancreatic ca. CAPAN2	36.3
Renal ca. UO-31	31.9	Pancreas Pool	11.2

Table MC. Panel 5 Islet

Tissue Name	Rel. Exp.(%) Ag6008, Run 245239907	Tissue Name	Rel. Exp.(%) Ag6008, Run 245239907
97457_Patient-02go_adipose	12.6	94709_Donor 2 AM - A_adipose	26.8
97476_Patient-07sk_skeletal muscle	9.5	94710_Donor 2 AM - B_adipose	26.4
97477_Patient-07ut_uterus	8.4	94711_Donor 2 AM - C_adipose	8.4
97478_Patient-07pl_placenta	16.4	94712_Donor 2 AD - A_adipose	37.6
99167_Bayer Patient 1	70.7	94713_Donor 2 AD - B_adipose	31.0
97482 Patient-	7.9	94714_Donor 2 AD - C_adipose	59.0

08ut_uterus			
97483_Patient-08pl_placenta	15.6	94742_Donor 3 U - A_Mesenchymal Stem Cells	11.0
97486_Patient-09sk_skeletal muscle	0.6	94743_Donor 3 U - B_Mesenchymal Stem Cells	34.2
97487_Patient-09ut_uterus	3.6	94730_Donor 3 AM - A_adipose	60.3
97488_Patient-09pl_placenta	9.6	94731_Donor 3 AM - B_adipose	27.4
97492_Patient-10ut_uterus	9.9	94732_Donor 3 AM - C_adipose	42.3
97493_Patient-10pl_placenta	18.3	94733_Donor 3 AD - A_adipose	100.0
97495_Patient-11go_adipose	5.5	94734_Donor 3 AD - B_adipose	44.1
97496_Patient-11sk_skeletal muscle	0.4	94735_Donor 3 AD - C_adipose	84.1
97497_Patient-11ut_uterus	3.5	77138_Liver_HepG2untreated	0.0
97498_Patient-11pl_placenta	11.0	73556_Heart_Cardiac stromal cells (primary)	14.8
97500_Patient-12go_adipose	7.4	81735_Small Intestine	9.5
97501_Patient-12sk_skeletal muscle	6.9	72409_Kidney_Proximal Convoluted Tubule	24.5
97502_Patient-12ut_uterus	9.3	82685_Small intestine_Duodenum	7.1
97503_Patient-12pl_placenta	6.1	90650_Adrenal_Adrenocortical adenoma	2.4
94721_Donor 2 U - A_Mesenchymal Stem Cells	6.7	72410_Kidney_HRCE	65.5
94722_Donor 2 U - B_Mesenchymal Stem Cells	13.6	72411_Kidney_HRE	46.0
94723_Donor 2 U - C_Mesenchymal Stem Cells	8.9	73139_Uterus_Uterine smooth muscle cells	30.4

**General\_screening\_panel\_v1.5 Summary:** Ag6008 Highest expression of this gene is detected in a lung cancer A549 cell line (CT=22.4). High expression of this gene is also seen in cluster of cancer cell lines derived from pancreatic, gastric, colon, lung, liver, renal, breast, ovarian, prostate, squamous cell carcinoma, melanoma and brain cancers. Thus,

5 expression of this gene could be used as a marker to detect the presence of these cancers.

Furthermore, therapeutic modulation of the expression or function of this gene may be effective in the treatment of pancreatic, gastric, colon, lung, liver, renal, breast, ovarian, prostate, squamous cell carcinoma, melanoma and brain cancers.

5 Among tissues with metabolic or endocrine function, this gene is expressed at high levels in pancreas, adipose, adrenal gland, thyroid, pituitary gland, skeletal muscle, heart, liver and the gastrointestinal tract. Therefore, therapeutic modulation of the activity of this gene through the use of small molecule drug may prove useful in the treatment of endocrine/metabolically related diseases, such as obesity and diabetes.

10 Interestingly, this gene is expressed at much higher levels in fetal (CTs=24-25), when compared to adult liver and lung (CTs=28-29). This observation suggests that expression of this gene can be used to distinguish fetal from adult lung and liver. In addition, the relative overexpression of this gene in fetal tissue suggests that the protein product may enhance lung and liver growth or development in the fetus and thus may also act in a regenerative capacity in the adult. Therefore, therapeutic modulation of the protein encoded by this gene  
15 could be useful in treatment of lung and liver related diseases.

In addition, this gene is expressed at high levels in all regions of the central nervous system examined, including amygdala, hippocampus, substantia nigra, thalamus, cerebellum, cerebral cortex, and spinal cord. Therefore, therapeutic modulation of this gene product may be useful in the treatment of central nervous system disorders such as Alzheimer's  
20 disease, Parkinson's disease, epilepsy, multiple sclerosis, schizophrenia and depression.

**Panel 5 Islet Summary:** Ag6008 Highest expression of this gene is detected in differentiated adipose (CT=27.7). This gene shows widespread expression in this panel. Moderate to high expression of this gene is detected in the tissues with metabolic/endocrine functions including islet cells, adipose, skeletal muscle, and gastrointestinal tracts.

25 This gene codes for ATP-citrate lyase. It is a major source of acetyl CoA that is the building block of lipid biosynthesis and provides substrate for the production of cholesterol. Reduced flux of acetyl CoA through the cholesterol biosynthetic pathway will prevent excess production of LXR alpha ligands. LXR alpha is a nuclear hormone receptor that is abundantly expressed in tissues associated with lipid metabolism. Activation of LXR  
30 alpha leads to the up-regulation of fatty acid synthesis. Thus, ATP-citrate lyase may be a

target for the treatment and/or prevention of obesity because its inhibition will decrease the availability of acetyl CoA for the synthesis of LXR alpha ligands, fatty acids, and triglycerides.

#### References:

- 5 1. Chawla A, Repa JJ, Evans RM, Mangelsdorf DJ. Nuclear receptors and lipid physiology: opening the X-files. *Science*. 2001 Nov 30;294(5548):1866-70. Review. PMID: 11729302.
2. Moon YA, Lee JJ, Park SW, Ahn YH, Kim KS. The roles of sterol regulatory element-binding proteins in the transactivation of the rat ATP citrate-lyase promoter. *J Biol Chem*. 2000 Sep 29;275(39):30280-6. PMID: 10801800.
- 10 3. Sato R, Okamoto A, Inoue J, Miyamoto W, Sakai Y, Emoto N, Shimano H, Maeda M. Transcriptional regulation of the ATP citrate-lyase gene by sterol regulatory element-binding proteins. *J Biol Chem*. 2000 Apr 28;275(17):12497-502. PMID: 10777536.

#### N. CG142631-01: serine dehydratase.

- 15 Expression of gene CG142631-01 was assessed using the primer-probe set Ag6006, described in Table NA. Results of the RTQ-PCR runs are shown in Tables NB, NC, ND and NE.

Table NA. Probe Name Ag6006

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-aagttcgtggatgatgagaaga-3'	22	858	510
Probe	TET-5'-ctggccgctgtctatagccacgt-3'-TAMRA	23	909	511
Reverse	5'-tccagttggagcttctggat-3'	20	933	512

Table NB. General\_screening\_panel\_v1.5

Tissue Name	Rel. Exp.(%) Ag6006, Run 228738305	Rel. Exp.(%) Ag6006, Run 228763464	Tissue Name	Rel. Exp.(%) Ag6006, Run 228738305	Rel. Exp.(%) Ag6006, Run 228763464
Adipose	2.8	3.1	Renal ca. TK-10	12.9	12.4
Melanoma*	0.0	0.0	Bladder	5.4	7.6

Hs688(A).T					
Melanoma* Hs688(B).T	0.0	0.0	Gastric ca. (liver met.) NCI-N87	1.1	0.9
Melanoma* M14	0.0	0.0	Gastric ca. KATO.III	0.0	0.0
Melanoma* LOXIMVI	0.0	0.0	Colon ca. SW- 948	0.0	0.0
Melanoma* SK-MEL-5	0.0	0.0	Colon ca. SW480	0.0	0.0
Squamous cell carcinoma SCC-4	0.0	0.0	Colon ca.* (SW480 met) SW620	0.0	0.0
Testis Pool	0.1	0.1	Colon ca. HT29	0.0	0.0
Prostate ca.* (bone met) PC-3	0.0	0.0	Colon ca. HCT- 116	0.0	0.0
Prostate Pool	0.2	0.1	Colon ca. CaCo- 2	0.1	0.0
Placenta	0.5	0.2	Colon cancer tissue	22.5	27.4
Uterus Pool	0.1	0.2	Colon ca. SW1116	0.0	0.0
Ovarian ca. OVCAR-3	0.7	0.3	Colon ca. Colo- 205	0.0	0.0
Ovarian ca. SK-OV-3	0.0	0.0	Colon ca. SW-48	0.0	0.0
Ovarian ca. OVCAR-4	0.0	0.0	Colon Pool	0.1	0.3
Ovarian ca. OVCAR-5	0.1	0.3	Small Intestine Pool	0.0	0.1
Ovarian ca. IGROV-1	0.0	0.0	Stomach Pool	1.5	1.2
Ovarian ca. OVCAR-8	0.1	0.0	Bone Marrow Pool	0.1	0.1
Ovary	0.6	0.6	Fetal Heart	0.0	0.0
Breast ca. MCF-7	0.0	0.0	Heart Pool	0.0	0.3
Breast ca. MDA-MB- 231	0.0	0.0	Lymph Node Pool	0.0	0.0
Breast ca. BT 549	0.0	0.1	Fetal Skeletal Muscle	0.0	0.0
Breast ca. T47D	0.0	0.0	Skeletal Muscle Pool	0.0	0.0

Breast ca. MDA-N	0.0	0.0	Spleen Pool	1.2	0.6
Breast Pool	0.3	0.0	Thymus Pool	0.2	0.0
Trachea	1.2	1.5	CNS cancer (glio/astro) U87- MG	0.0	0.0
Lung	0.0	0.0	CNS cancer (glio/astro) U- 118-MG	0.1	0.0
Fetal Lung	0.9	1.8	CNS cancer (neuro;met) SK- N-AS	0.0	0.0
Lung ca. NCI-N417	0.0	0.0	CNS cancer (astro) SF-539	0.2	0.0
Lung ca. LX- 1	0.0	0.0	CNS cancer (astro) SNB-75	0.1	0.0
Lung ca. NCI-H146	0.0	0.0	CNS cancer (glio) SNB-19	0.0	0.0
Lung ca. SHP-77	0.1	0.0	CNS cancer (glio) SF-295	0.0	0.2
Lung ca. A549	1.7	1.4	Brain (Amygdala) Pool	3.8	2.9
Lung ca. NCI-H526	0.0	0.0	Brain (cerebellum)	7.9	10.2
Lung ca. NCI-H23	0.0	0.0	Brain (fetal)	0.5	0.6
Lung ca. NCI-H460	0.0	0.0	Brain (Hippocampus) Pool	3.7	5.9
Lung ca. HOP-62	0.0	0.0	Cerebral Cortex Pool	2.2	2.4
Lung ca. NCI-H522	0.0	0.1	Brain (Substantia nigra) Pool	3.1	3.3
Liver	<b>100.0</b>	<b>100.0</b>	Brain (Thalamus) Pool	3.4	3.5
Fetal Liver	0.9	0.8	Brain (whole)	4.8	3.2
Liver ca. HepG2	0.0	0.0	Spinal Cord Pool	2.0	1.8
Kidney Pool	0.1	0.1	Adrenal Gland	13.2	12.7
Fetal Kidney	0.0	0.0	Pituitary gland Pool	0.0	0.0
Renal ca. 786-0	0.2	0.1	Salivary Gland	0.2	0.2
Renal ca. A498	0.0	0.1	Thyroid (female)	0.4	0.7

Renal ca. ACHN	0.0	0.0	Pancreatic ca. CAPAN2	0.0	0.0
Renal ca. UO-31	0.0	0.0	Pancreas Pool	0.3	0.3

Table NC. Oncology\_cell\_line\_screening\_panel\_v3.1

Tissue Name	Rel. Exp.(%) Ag6006, Run 22513897 6	Rel. Exp.(%) Ag6006, Run 23027712 9	Tissue Name	Rel. Exp.(%) Ag6006, Run 225138976	Rel. Exp.(%) Ag6006, Run 230277129
Daoy Medulloblastoma/Cerebellum	0.0	0.0	Ca Ski_Cervical epidermoid carcinoma (metastasis)	0.0	0.0
TE671 Medulloblastom/Cerebellum	0.0	0.0	ES-2_Ovarian clear cell carcinoma	0.0	0.0
D283 Med Medulloblastoma/Cerebellum	0.0	0.0	Ramos/6h stim_ Stimulated with PMA/ionomycin 6h	0.0	0.0
PFSK-1 Primitive Neuroectodermal/Cerebellum	13.3	3.1	Ramos/14h stim_ Stimulated with PMA/ionomycin 14h	0.0	0.0
XF-498_CNS	0.0	0.0	MEG-01_Chronic myelogenous leukemia (megokaryoblast)	2.2	6.9
SNB-78_CNS/glioma	0.0	0.0	Raji_Burkitt's lymphoma	0.0	0.0
SF-268_CNS/glioblastoma	0.0	0.0	Daudi_Burkitt's lymphoma	0.0	0.0
T98G_Glioblastoma	0.0	0.0	U266_B-cell plasmacytoma/myelo ma	0.0	3.8
SK-N-SH_Neuroblastoma (metastasis)	0.0	0.0	CA46_Burkitt's lymphoma	0.0	0.0
SF-295_CNS/glioblastoma	0.0	0.0	RL_non-Hodgkin's B-cell lymphoma	0.0	0.0
Cerebellum	66.9	97.9	JM1_pre-B-cell lymphoma/leukemia	0.0	0.0
Cerebellum	100.0	100.0	Jurkat_T cell leukemia	0.0	0.0
NCI-H292_Mucoepidermoid lung ca.	0.0	0.0	TF- 1 Erythroleukemia	12.2	10.4

DMS-114_Small cell lung cancer	0.0	0.0	HUT 78_T-cell lymphoma	0.0	0.0
DMS-79_Small cell lung cancer/neuroendocrine	0.0	0.0	U937_Histiocytic lymphoma	43.5	42.3
NCI-H146_Small cell lung cancer/neuroendocrine	0.0	0.0	KU-812_Myelogenous leukemia	2.3	0.0
NCI-H526_Small cell lung cancer/neuroendocrine	0.0	0.0	769-P_Clear cell renal ca.	0.0	0.0
NCI-N417_Small cell lung cancer/neuroendocrine	0.0	0.0	Caki-2_Clear cell renal ca.	0.0	0.0
NCI-H82_Small cell lung cancer/neuroendocrine	3.7	0.0	SW 839_Clear cell renal ca.	0.0	0.0
NCI-H157_Squamous cell lung cancer (metastasis)	0.0	0.0	G401_Wilms' tumor	8.3	20.7
NCI-H1155_Large cell lung cancer/neuroendocrine	0.0	0.0	Hs766T_Pancreatic ca. (LN metastasis)	2.0	0.0
NCI-H1299_Large cell lung cancer/neuroendocrine	0.0	0.0	CAPAN-1_Pancreatic adenocarcinoma (liver metastasis)	0.0	0.0
NCI-H727_Lung carcinoid	0.0	0.0	SU86.86_Pancreatic carcinoma (liver metastasis)	0.0	0.0
NCI-UMC-11_Lung carcinoid	0.0	0.0	BxPC-3_Pancreatic adenocarcinoma	0.0	0.0
LX-1_Small cell lung cancer	0.0	0.0	HPAC_Pancreatic adenocarcinoma	0.0	0.0
Colo-205_Colon cancer	0.0	0.0	MIA PaCa-2_Pancreatic ca.	0.0	0.0
KM12_Colon cancer	0.0	0.0	CFPAC-1_Pancreatic ductal adenocarcinoma	0.6	0.0
KM20L2_Colon cancer	0.0	0.0	PANC-1_Pancreatic epithelioid ductal ca.	0.0	0.0
NCI-H716_Colon cancer	0.0	0.0	T24_Bladder ca. (transitional cell)	0.0	0.0
SW-48_Colon adenocarcinoma	0.0	0.0	5637_Bladder ca.	0.0	0.0
SW1116_Colon adenocarcinoma	0.0	0.0	HT-1197_Bladder ca.	2.3	0.0
LS 174T_Colon adenocarcinoma	0.0	0.0	UM-UC-3_Bladder ca. (transitional cell)	0.0	0.0
SW-948_Colon adenocarcinoma	0.0	0.0	A204_Rhabdomyosarcoma	0.0	0.0

SW-480_Colon adenocarcinoma	0.0	0.0	HT-1080_Fibrosarcoma	0.0	2.0
NCI-SNU-5_Gastric ca.	0.0	0.0	MG-63_Osteosarcoma (bone)	0.0	8.0
KATO III_Stomach	0.5	0.0	SK-LMS-1_Leiomyosarcoma (vulva)	3.7	0.0
NCI-SNU-16_Gastric ca.	2.6	0.0	SJRH30_Rhabdomyosarcoma (met to bone marrow)	0.0	0.0
NCI-SNU-1_Gastric ca.	0.0	0.0	A431_Epidermoid ca.	1.5	0.0
RF-1_Gastric adenocarcinoma	7.4	11.3	WM266-4_Melanoma	1.6	3.8
RF-48_Gastric adenocarcinoma	17.1	7.8	DU 145_Prostate	0.0	0.0
MKN-45_Gastric ca.	0.0	0.0	MDA-MB-468_Breast adenocarcinoma	2.4	0.0
NCI-N87_Gastric ca.	0.0	0.0	SSC-4_Tongue	0.0	0.0
OVCAR-5_Ovarian ca.	0.0	0.0	SSC-9_Tongue	0.0	0.0
RL95-2_Uterine carcinoma	2.0	0.0	SSC-15_Tongue	0.0	0.0
HelaS3_Cervical adenocarcinoma	0.0	0.0	CAL 27_Squamous cell ca. of tongue	0.0	0.0

Table ND. Panel 4.1D

Tissue Name	Rel. Exp.(%) Ag6006, Run 225787022	Tissue Name	Rel. Exp.(%) Ag6006, Run 225787022
Secondary Th1 act	0.0	HUVEC IL-1beta	0.0
Secondary Th2 act	0.0	HUVEC IFN gamma	0.0
Secondary Tr1 act	0.0	HUVEC TNF alpha + IFN gamma	0.0
Secondary Th1 rest	0.2	HUVEC TNF alpha + IL4	0.0
Secondary Th2 rest	0.0	HUVEC IL-11	0.4
Secondary Tr1 rest	0.0	Lung Microvascular EC none	0.6
Primary Th1 act	0.0	Lung Microvascular EC TNFalpha + IL-1beta	0.0
Primary Th2 act	0.0	Microvascular Dermal EC none	0.0
Primary Tr1 act	0.0	Microvascular Dermal EC TNFalpha + IL-1beta	0.0

Primary Th1 rest	0.0	Bronchial epithelium TNFalpha + IL1beta	0.0
Primary Th2 rest	0.0	Small airway epithelium none	0.0
Primary Tr1 rest	0.0	Small airway epithelium TNFalpha + IL-1beta	0.0
CD45RA CD4 lymphocyte act	0.0	Coronary artery SMC rest	0.0
CD45RO CD4 lymphocyte act	0.0	Coronary artery SMC TNFalpha + IL-1beta	0.0
CD8 lymphocyte act	0.0	Astrocytes rest	0.0
Secondary CD8 lymphocyte rest	0.0	Astrocytes TNFalpha + IL-1beta	0.0
Secondary CD8 lymphocyte act	0.0	KU-812 (Basophil) rest	0.2
CD4 lymphocyte none	0.0	KU-812 (Basophil) PMA/ionomycin	0.2
2ry Th1/Th2/Tr1 _anti- CD95 CH11	0.0	CCD1106 (Keratinocytes) none	0.0
LAK cells rest	7.5	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	0.0
LAK cells IL-2	0.0	Liver cirrhosis	100.0
LAK cells IL-2+IL-12	0.2	NCI-H292 none	0.0
LAK cells IL-2+IFN gamma	0.2	NCI-H292 IL-4	0.0
LAK cells IL-2+ IL-18	0.0	NCI-H292 IL-9	0.0
LAK cells PMA/ionomycin	3.6	NCI-H292 IL-13	0.0
NK Cells IL-2 rest	0.0	NCI-H292 IFN gamma	0.0
Two Way MLR 3 day	1.3	HPAEC none	0.0
Two Way MLR 5 day	1.3	HPAEC TNF alpha + IL- 1 beta	0.0
Two Way MLR 7 day	1.1	Lung fibroblast none	0.0
PBMC rest	0.5	Lung fibroblast TNF alpha + IL-1 beta	0.2
PBMC PWM	0.0	Lung fibroblast IL-4	0.3
PBMC PHA-L	0.0	Lung fibroblast IL-9	0.0
Ramos (B cell) none	0.0	Lung fibroblast IL-13	0.0
Ramos (B cell) ionomycin	0.0	Lung fibroblast IFN gamma	0.0
B lymphocytes PWM	0.0	Dermal fibroblast CCD1070 rest	0.0
B lymphocytes CD40L	0.0	Dermal fibroblast	0.0

and IL-4		CCD1070 TNF alpha	
EOL-1 dbcAMP	0.0	Dermal fibroblast CCD1070 IL-1 beta	0.0
EOL-1 dbcAMP PMA/ionomycin	0.0	Dermal fibroblast IFN gamma	0.0
Dendritic cells none	8.1	Dermal fibroblast IL-4	0.0
Dendritic cells LPS	10.4	Dermal Fibroblasts rest	0.0
Dendritic cells anti- CD40	7.1	Neutrophils TNFa+LPS	0.0
Monocytes rest	0.4	Neutrophils rest	0.0
Monocytes LPS	16.0	Colon	0.1
Macrophages rest	87.7	Lung	1.2
Macrophages LPS	82.4	Thymus	3.2
HUVEC none	0.0	Kidney	2.5
HUVEC starved	0.0		

Table NE. Panel 5 Islet

Tissue Name	Rel. Exp.(%) Ag6006, Run 22505116 4	Rel. Exp.(%) Ag6006, Run 24898915 2	Rel. Exp.(%) Ag6006, Run 24913905 5	Tissue Name	Rel. Exp.(%) Ag6006, Run 22505116 4	Rel. Exp.(%) Ag6006, Run 248989152 2	Rel. Exp.(%) Ag6006, Run 249139055 5
97457_Patient- 02go adipose	6.5	0.0	20.0	94709_Donor 2 AM - A adipose	0.0	0.0	0.0
97476_Patient- 07sk skeletal muscle	20.7	0.0	15.6	94710_Donor 2 AM - B adipose	0.0	0.0	0.0
97477_Patient- 07ut uterus	6.7	0.0	0.0	94711_Donor 2 AM - C adipose	0.0	0.0	0.0
97478_Patient- 07pl placenta	11.8	0.0	5.0	94712_Donor 2 AD - A adipose	0.0	0.0	0.0
99167_Bayer Patient 1	88.3	100.0	62.0	94713_Donor 2 AD - B adipose	0.0	0.0	0.0
97482_Patient- 08ut uterus	8.5	6.7	0.0	94714_Donor 2 AD - C adipose	0.0	0.0	0.0
97483_Patient- 08pl placenta	4.4	13.5	5.4	94742_Donor 3 U - A_Mesenchymal Stem Cells	0.0	0.0	0.0
97486_Patient- 09sk skeletal muscle	0.0	0.0	0.0	94743_Donor 3 U - B_Mesenchymal Stem Cells	0.0	0.0	0.0

97487_Patient-09ut_uterus	0.0	0.0	0.0	94730_Donor 3 AM - A_adipose	0.0	0.0	0.0
97488_Patient-09pl_placenta	4.9	0.0	0.0	94731_Donor 3 AM - B_adipose	0.0	0.0	0.0
97492_Patient-10ut_uterus	0.0	0.0	0.0	94732_Donor 3 AM - C_adipose	0.0	0.0	0.0
97493_Patient-10pl_placenta	4.6	0.0	5.1	94733_Donor 3 AD - A_adipose	0.0	0.0	0.0
97495_Patient-11go_adipose	0.0	0.0	3.8	94734_Donor 3 AD - B_adipose	0.0	0.0	0.0
97496_Patient-11sk_skeletal muscle	0.0	0.0	0.0	94735_Donor 3 AD - C_adipose	0.0	0.0	0.0
97497_Patient-11ut_uterus	0.0	0.0	0.0	77138_Liver_He pG2untreated	0.0	0.0	0.0
97498_Patient-11pl_placenta	0.0	0.0	0.0	73556_Heart_Cardiac stromal cells (primary)	0.0	0.0	0.0
97500_Patient-12go_adipose	0.0	6.0	4.9	81735_Small Intestine	8.5	6.3	5.1
97501_Patient-12sk_skeletal muscle	4.0	0.0	9.2	72409_Kidney_Proximal Convoluted Tubule	0.0	0.0	0.0
97502_Patient-12ut_uterus	0.0	5.1	0.0	82685_Small intestine_Duodenum	0.0	0.0	5.4
97503_Patient-12pl_placenta	14.9	7.3	7.7	90650_Adrenal Adrenocortical adenoma	100.0	49.3	100.0
94721_Donor 2 U - A_Mesenchymal Stem Cells	0.0	0.0	0.0	72410_Kidney_HRCE	0.0	0.0	0.0
94722_Donor 2 U - B_Mesenchymal Stem Cells	0.0	0.0	0.0	72411_Kidney_HRE	0.0	0.0	0.0
94723_Donor 2 U - C_Mesenchymal Stem Cells	0.0	0.0	2.4	73139_Uterus_Uterine smooth muscle cells	0.0	0.0	0.0

**General\_screening\_panel\_v1.5 Summary:** Ag6006 Two experiments with same probe-primer sets are in excellent agreement with highest expression of this gene detected in liver (CTs=26). Interestingly, expression of this gene is higher in adult as compared to fetal liver

(CTs=32-33). Therefore, expression of this gene may be useful in distinguishing between adult and fetal liver.

- In addition, moderate to low expression of this gene is also detected in tissues with metabolic/endocrine functions including pancreas, adipose, adrenal gland, thyroid, and stomach. This gene codes for Serine dehydratase (SD). SD catalyzes the PLP-dependent alpha, beta-elimination of L-serine to pyruvate and ammonia. It is one of three enzymes that are regarded as metabolic exits of the serine-glycine pool. SD is critical for hepatic glucose production. Therefore, inhibition of SD would decrease gluconeogenesis, thus an antagonist of SD would be beneficial for treatment hyperglycemia and diabetes.
- 10 In addition moderate levels of expression of this gene is in all regions of the central nervous system examined, including amygdala, hippocampus, substantia nigra, thalamus, cerebellum, cerebral cortex, and spinal cord. Therefore, therapeutic modulation of this gene product may be useful in the treatment of central nervous system disorders such as Alzheimer's disease, Parkinson's disease, epilepsy, multiple sclerosis, schizophrenia and depression.
- 15

- Oncology\_cell\_line\_screening\_panel\_v3.1 Summary:** Ag6006 Two experiments with same probe-primer sets are in excellent agreement, with highest expression of this gene detected in cerebellum (CTs=32-33.7). In addition, low levels of expression of this gene is also detected in histiocytic lymphoma. Therefore, therapeutic modulation of this gene may be useful in the treatment of ataxia, autism and histiocytic lymphoma.
- 20

- Panel 4.1D Summary:** Ag6006 Highest expression of this gene is detected in liver cirrhosis sample (CT=29). In addition, moderate to low expression of this gene resting macrophage, LPS activated monocytes and macrophages, dendritic cells, resting and PMA/ionomycin activated LAK cells and normal tissues represented by thymus and kidney. Therefore, therapeutic modulation of this gene may be useful in the treatment of liver cirrhosis, asthma, emphysema, inflammatory bowel disease, arthritis and psoriasis.
- 25

Results from another experiment with this gene (run 225245206) are not included. The amp plot indicates that there were experimental difficulties with this run.

**Panel 5 Islet Summary:** Ag6006 Three experiments with same probe and primer sets are in good agreement. Low expression of this gene is detected mainly in islet cells and adrenocortical adenoma cells (CTs=33-34.8). Therefore, therapeutic modulation of this gene of SD encoded by this gene through the use of small molecule drug may be useful in the treatment of adrenocortical adenoma and metabolic disorders especially type II diabetes.

#### O. CG151359-01: LACTATE DEHYDROGENASE A Like.

Expression of gene CG151359-01 was assessed using the primer-probe set Ag5225, described in Table OA. Results of the RTQ-PCR runs are shown in Table OB.

10 Table OA. Probe Name Ag5225

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-tggtattggaagcggctgta-3'	20	618	513
Probe	TET-5'-ctgttcgttttcaattcttcattgga-3'-TAMRA	26	647	514
Reverse	5'-cagagtggataccaagcttttg-3'	22	673	515

Table OB. General\_screening\_panel\_v1.5

Tissue Name	Rel. Exp.(%) Ag5225, Run 228763462	Tissue Name	Rel. Exp.(%) Ag5225, Run 228763462
Adipose	0.0	Renal ca. TK-10	0.0
Melanoma* Hs688(A).T	0.0	Bladder	0.7
Melanoma* Hs688(B).T	0.0	Gastric ca. (liver met.) NCI-N87	0.0
Melanoma* M14	0.0	Gastric ca. KATO III	0.0
Melanoma* LOXIMVI	7.9	Colon ca. SW-948	0.0
Melanoma* SK- MEL-5	0.0	Colon ca. SW480	0.0
Squamous cell carcinoma SCC-4	0.0	Colon ca.* (SW480 met) SW620	0.0
Testis Pool	100.0	Colon ca. HT29	0.6
Prostate ca.* (bone met) PC-3	0.0	Colon ca. HCT-116	0.0
Prostate Pool	10.5	Colon ca. CaCo-2	49.0
Placenta	0.0	Colon cancer tissue	0.0

Uterus Pool	0.0	Colon ca. SW1116	0.0
Ovarian ca. OVCAR-3	0.0	Colon ca. Colo-205	0.0
Ovarian ca. SK-OV-3	0.0	Colon ca. SW-48	0.0
Ovarian ca. OVCAR-4	3.7	Colon Pool	75.3
Ovarian ca. OVCAR-5	0.0	Small Intestine Pool	1.8
Ovarian ca. IGROV-1	0.0	Stomach Pool	0.0
Ovarian ca. OVCAR-8	0.0	Bone Marrow Pool	0.0
Ovary	0.0	Fetal Heart	0.0
Breast ca. MCF-7	0.0	Heart Pool	0.0
Breast ca. MDA-MB-231	0.0	Lymph Node Pool	0.0
Breast ca. BT 549	0.0	Fetal Skeletal Muscle	0.0
Breast ca. T47D	0.0	Skeletal Muscle Pool	0.0
Breast ca. MDA-N	0.0	Spleen Pool	0.0
Breast Pool	0.0	Thymus Pool	0.0
Trachea	0.0	CNS cancer (glio/astro) U87-MG	0.0
Lung	0.0	CNS cancer (glio/astro) U-118-MG	25.2
Fetal Lung	0.0	CNS cancer (neuro;met) SK-N-AS	0.0
Lung ca. NCI-N417	0.0	CNS cancer (astro) SF-539	0.0
Lung ca. LX-1	0.0	CNS cancer (astro) SNB-75	0.0
Lung ca. NCI-H146	0.0	CNS cancer (glio) SNB-19	0.0
Lung ca. SHP-77	0.0	CNS cancer (glio) SF-295	0.0
Lung ca. A549	0.0	Brain (Amygdala) Pool	0.0
Lung ca. NCI-H526	0.0	Brain (cerebellum)	0.0
Lung ca. NCI-H23	0.0	Brain (fetal)	0.0
Lung ca. NCI-H460	0.0	Brain (Hippocampus) Pool	0.0
Lung ca. HOP-62	0.0	Cerebral Cortex Pool	0.0
Lung ca. NCI-H522	0.0	Brain (Substantia nigra) Pool	8.6
Liver	0.0	Brain (Thalamus) Pool	0.0

Fetal Liver	5.8	Brain (whole)	0.0
Liver ca. HepG2	0.0	Spinal Cord Pool	0.0
Kidney Pool	0.3	Adrenal Gland	0.0
Fetal Kidney	0.0	Pituitary gland Pool	0.0
Renal ca. 786-0	0.0	Salivary Gland	0.0
Renal ca. A498	0.0	Thyroid (female)	0.0
Renal ca. ACHN	0.0	Pancreatic ca. CAPAN2	0.0
Renal ca. UO-31	0.0	Pancreas Pool	0.0

**CNS\_neurodegeneration\_v1.0 Summary:** Ag5225 Expression of this gene is low/undetectable in all samples on this panel (CTs>35). (Data not shown.)

- General\_screening\_panel\_v1.5 Summary:** Ag5225 Expression of this gene is limited to a few samples on this panel, with highest expression seen in testis (CT=31.8). Moderate to low levels of expression are also seen in normal colon, a colon cancer cell line, and a brain cancer cell line.

**Panel 4.1D Summary:** Ag5225 Expression of this gene is low/undetectable in all samples on this panel (CTs>35). (Data not shown.)

- Panel 5 Islet Summary:** Ag5225 Expression of this gene is low/undetectable in all samples on this panel (CTs>35). (Data not shown.)

**P. CG152227-01: 3-HYDROXYISOBUTYRYL-COENZYME A HYDROLASE.**

Expression of gene CG152227-01 was assessed using the primer-probe set Ag6857, described in Table PA.

- 15 Table PA. Probe Name Ag6857

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-ttggactctggtcttcaagtat-3'	22	186	516
Probe	TET-5'- agacttgtctcgatcaatcttagactctgtatggtaa-3'- TAMRA	37	211	517
Reverse	5'-cttcaaaagaaaatattgcacatcg-3'	24	258	518

**General\_screening\_panel\_v1.6 Summary:** Ag6857 Expression of this gene is low/undetectable in all samples on this panel (CTs>35). (Data not shown.)

**Q. CG152547-01: Similar to Zinc transporter 1.**

- 5 Expression of gene CG152547-01 was assessed using the primer-probe set Ag7619, described in Table QA.

Table QA. Probe Name Ag7619

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-tgctcatcttccatcaccaa-3'	20	392	519
Probe	TET-5'-ccctaattctcaagtaatcaggacacaa-3'- TAMRA	28	413	520
Reverse	5'-tggttttcctaggcagagga-3'	20	462	521

**CNS\_neurodegeneration\_v1.0 Summary:** Ag7619 Expression of this gene is low/undetectable in all samples on this panel (CTs>35). (Data not shown.)

- 10 **Panel 4.1D Summary:** Ag7619 Expression of this gene is low/undetectable in all samples on this panel (CTs>35). (Data not shown.)

**R. CG152646-01: Amidase.**

Expression of gene CG152646-01 was assessed using the primer-probe set Ag6876, described in Table RA.

- 15 Table RA. Probe Name Ag6876

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-cacatctgtgaccatattggt-3'	21	573	522
Probe	TET-5'-tttaactgggtccaaatacaccatctgtg-3'- TAMRA	28	613	523
Reverse	5'-tttgctatgggatctg-3'	16	645	524

**General\_screening\_panel\_v1.6 Summary:** Ag6876 Expression of this gene is low/undetectable in all samples on this panel (CTs>35). (Data not shown.)

**S. CG152959-01: Prenyl protein-specific endoprotease 2.**

Expression of gene CG152959-01 was assessed using the primer-probe set Ag7172, described in Table SA. Results of the RTQ-PCR runs are shown in Table SB. Please note that CG152959-01 represents a full-length physical clone.

Table SA. Probe Name Ag7172

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5' - cctggaggacgtgctgt-3'	17	191	525
Probe	TET-5' -ccaacctgtcagagtggctgagtccc-3' - TAMRA	26	223	526
Reverse	5' -gcgcttgcggaagg-3'	14	273	527

5 Table SB. General\_screening\_panel\_v1.7

Tissue Name	Rel. Exp.(%) Ag7172, Run 318039790	Tissue Name	Rel. Exp.(%) Ag7172, Run 318039790
Adipose	10.6	Gastric ca. (liver met.) NCI-N87	0.0
HUVEC	35.8	Stomach	0.0
Melanoma* Hs688(A).T	0.3	Colon ca. SW-948	6.0
Melanoma* Hs688(B).T	66.9	Colon ca. SW480	0.4
Melanoma (met) SK-MEL-5	4.4	Colon ca. (SW480 met) SW620	6.8
Testis	13.5	Colon ca. HT29	30.4
Prostate ca. (bone met) PC-3	0.5	Colon ca. HCT-116	22.2
Prostate ca. DU145	19.3	Colon cancer tissue	1.0
Prostate pool	7.7	Colon ca. SW1116	6.1
Uterus pool	2.5	Colon ca. Colo-205	11.0
Ovarian ca. OVCAR-3	14.1	Colon ca. SW-48	9.4
Ovarian ca. (ascites) SK-OV-3	0.8	Colon	15.9
Ovarian ca. OVCAR-4	51.4	Small Intestine	1.5
Ovarian ca. OVCAR-5	29.1	Fetal Heart	0.7
Ovarian ca. IGROV-1	100.0	Heart	1.2
Ovarian ca. OVCAR-8	24.0	Lymph Node pool	3.1

Ovary	3.2	Lymph Node pool	26.1
Breast ca. MCF-7	17.7	Fetal Skeletal Muscle	1.7
Breast ca. MDA-MB-231	43.8	Skeletal Muscle pool	0.3
Breast ca. BT-549	14.1	Skeletal Muscle	0.2
Breast ca. T47D	15.5	Spleen	4.4
Breast pool	7.5	Thymus	14.7
Trachea	15.8	CNS cancer (glio/astro) SF-268	6.4
Lung	1.2	CNS cancer (glio/astro) T98G	3.3
Fetal Lung	9.0	CNS cancer (neuro;met) SK-N-AS	0.2
Lung ca. NCI-N417	10.0	CNS cancer (astro) SF-539	8.9
Lung ca. LX-1	4.4	CNS cancer (astro) SNB-75	10.1
Lung ca. NCI-H146	15.5	CNS cancer (glio) SNB-19	16.5
Lung ca. SHP-77	38.2	CNS cancer (glio) SF-295	4.9
Lung ca. NCI-H23	26.2	Brain (Amygdala)	6.6
Lung ca. NCI-H460	8.5	Brain (Cerebellum)	12.8
Lung ca. HOP-62	9.6	Brain (Fetal)	25.5
Lung ca. NCI-H522	56.3	Brain (Hippocampus)	4.7
Lung ca. DMS-114	8.8	Cerebral Cortex pool	1.8
Liver	0.0	Brain (Substantia nigra)	4.0
Fetal Liver	1.0	Brain (Thalamus)	4.3
Kidney pool	32.3	Brain (Whole)	21.6
Fetal Kidney	3.7	Spinal Cord	0.8
Renal ca. 786-0	40.1	Adrenal Gland	2.2
Renal ca. A498	12.7	Pituitary Gland	11.9
Renal ca. ACHN	15.0	Salivary Gland	8.0
Renal ca. UO-31	22.8	Thyroid	8.4
Renal ca. TK-10	46.0	Pancreatic ca. PANC-1	10.5
Bladder	1.6	Pancreas pool	1.5

**General\_screening\_panel\_v1.7 Summary:** Ag7172 Highest expression of this gene is detected in ovarian cancer IGROV-1 cell line (CT=28.3). Moderate levels of expression of this gene is also seen in cluster of cancer cell lines derived from pancreatic, gastric, colon,

lung, liver, renal, breast, ovarian, melanoma and brain cancers. Thus, expression of this gene could be used as a marker to detect the presence of these cancers. Furthermore, therapeutic modulation of the expression or function of this gene may be effective in the treatment of pancreatic, gastric, colon, lung, liver, renal, breast, ovarian, melanoma and

5 brain cancers.

Among tissues with metabolic or endocrine function, this gene is expressed at moderate to low levels in pancreas, adipose, adrenal gland, thyroid, pituitary gland, fetal skeletal muscle, heart, fetal liver and the gastrointestinal tract. Therefore, therapeutic modulation of the activity of this gene may prove useful in the treatment of endocrine/metabolically

10 related diseases, such as obesity and diabetes.

In addition, this gene is expressed at low levels in all regions of the central nervous system examined, including amygdala, hippocampus, substantia nigra, thalamus, cerebellum, cerebral cortex, and spinal cord. Therefore, therapeutic modulation of this gene product may be useful in the treatment of central nervous system disorders such as Alzheimer's

15 disease, Parkinson's disease, epilepsy, multiple sclerosis, schizophrenia and depression.

#### **T. CG153033-01: NA-DEPENDENT INORGANIC PHOSPHATE COTRANSPORTER.**

Expression of gene CG153033-01 was assessed using the primer-probe set Ag5798, described in Table TA. Results of the RTQ-PCR runs are shown in Tables TB and TC.

20 Table TA. Probe Name Ag5798

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5' -aatcttggagttgccattgtg-3'	21	223	528
Probe	TET-5' -ccatcaacatatacgggtgctattgttgacc-3' -TAMRA	30	249	529
Reverse	5' -tcccagttaaactgtgctgtct-3'	22	284	530

Table TB. CNS\_neurodegeneration\_v1.0

Tissue Name	Rel. Exp.(%) Ag5798, Run 247179626	Tissue Name	Rel. Exp.(%) Ag5798, Run 247179626
AD 1 Hippo	8.0	Control (Path) 3 Temporal Ctx	0.0
AD 2 Hippo	14.4	Control (Path) 4	39.0

		Temporal Ctx	
AD 3 Hippo	3.7	AD 1 Occipital Ctx	0.0
AD 4 Hippo	7.3	AD 2 Occipital Ctx (Missing)	0.0
AD 5 hippo	24.1	AD 3 Occipital Ctx	0.0
AD 6 Hippo	24.8	AD 4 Occipital Ctx	24.7
Control 2 Hippo	42.6	AD 5 Occipital Ctx	9.3
Control 4 Hippo	3.3	AD 6 Occipital Ctx	40.6
Control (Path) 3 Hippo	0.0	Control 1 Occipital Ctx	3.0
AD 1 Temporal Ctx	9.3	Control 2 Occipital Ctx	21.3
AD 2 Temporal Ctx	94.6	Control 3 Occipital Ctx	3.5
AD 3 Temporal Ctx	3.6	Control 4 Occipital Ctx	0.0
AD 4 Temporal Ctx	13.6	Control (Path) 1 Occipital Ctx	54.0
AD 5 Inf Temporal Ctx	100.0	Control (Path) 2 Occipital Ctx	0.0
AD 5 Sup Temporal Ctx	71.7	Control (Path) 3 Occipital Ctx	0.0
AD 6 Inf Temporal Ctx	57.8	Control (Path) 4 Occipital Ctx	3.4
AD 6 Sup Temporal Ctx	22.8	Control 1 Parietal Ctx	0.0
Control 1 Temporal Ctx	0.0	Control 2 Parietal Ctx	59.9
Control 2 Temporal Ctx	38.7	Control 3 Parietal Ctx	0.0
Control 3 Temporal Ctx	12.6	Control (Path) 1 Parietal Ctx	46.7
Control 4 Temporal Ctx	10.0	Control (Path) 2 Parietal Ctx	16.0
Control (Path) 1 Temporal Ctx	70.2	Control (Path) 3 Parietal Ctx	7.8
Control (Path) 2 Temporal Ctx	8.2	Control (Path) 4 Parietal Ctx	17.6

Table TC. General\_screening\_panel\_v1.5

Tissue Name	Rel. Exp.(%) Ag5798, Run 245274436	Tissue Name	Rel. Exp.(%) Ag5798, Run 245274436
Adipose	0.0	Renal ca. TK-10	0.0
Melanoma* Hs688(A).T	0.0	Bladder	1.5
Melanoma* Hs688(B).T	0.0	Gastric ca. (liver met.) NCI-N87	1.0
Melanoma* M14	0.0	Gastric ca. KATO III	0.0
Melanoma* LOXIMVI	1.9	Colon ca. SW-948	0.0
Melanoma* SK- MEL-5	0.0	Colon ca. SW480	0.0
Squamous cell carcinoma SCC-4	0.9	Colon ca.* (SW480 met) SW620	0.0
Testis Pool	5.9	Colon ca. HT29	0.0
Prostate ca.* (bone met) PC-3	0.4	Colon ca. HCT-116	0.0
Prostate Pool	0.0	Colon ca. CaCo-2	0.0
Placenta	3.9	Colon cancer tissue	0.9
Uterus Pool	0.0	Colon ca. SW1116	0.0
Ovarian ca. OVCAR-3	0.0	Colon ca. Colo-205	0.0
Ovarian ca. SK- OV-3	0.0	Colon ca. SW-48	0.0
Ovarian ca. OVCAR-4	0.0	Colon Pool	0.0
Ovarian ca. OVCAR-5	0.0	Small Intestine Pool	2.1
Ovarian ca. IGROV-1	1.0	Stomach Pool	1.3
Ovarian ca. OVCAR-8	0.0	Bone Marrow Pool	1.6
Ovary	3.1	Fetal Heart	5.9
Breast ca. MCF-7	0.0	Heart Pool	0.0
Breast ca. MDA- MB-231	0.0	Lymph Node Pool	0.0
Breast ca. BT 549	0.0	Fetal Skeletal Muscle	0.0
Breast ca. T47D	0.0	Skeletal Muscle Pool	1.1
Breast ca. MDA-N	0.0	Spleen Pool	0.9
Breast Pool	1.6	Thymus Pool	7.0
Trachea	0.0	CNS cancer (glio/astro) U87-MG	0.0
Lung	0.0	CNS cancer	0.0

		(glio/astro) U-118-MG	
Fetal Lung	27.9	CNS cancer (neuro;met) SK-N-AS	0.0
Lung ca. NCI-N417	0.0	CNS cancer (astro) SF-539	0.0
Lung ca. LX-1	1.2	CNS cancer (astro) SNB-75	0.0
Lung ca. NCI-H146	0.0	CNS cancer (glio) SNB-19	0.0
Lung ca. SHP-77	0.0	CNS cancer (glio) SF-295	0.0
Lung ca. A549	0.8	Brain (Amygdala) Pool	14.7
Lung ca. NCI-H526	90.8	Brain (cerebellum)	3.4
Lung ca. NCI-H23	0.6	Brain (fetal)	20.2
Lung ca. NCI-H460	0.0	Brain (Hippocampus) Pool	45.4
Lung ca. HOP-62	0.0	Cerebral Cortex Pool	19.6
Lung ca. NCI-H522	0.0	Brain (Substantia nigra) Pool	29.7
Liver	1.6	Brain (Thalamus) Pool	100.0
Fetal Liver	51.1	Brain (whole)	10.4
Liver ca. HepG2	0.0	Spinal Cord Pool	6.1
Kidney Pool	0.0	Adrenal Gland	0.0
Fetal Kidney	5.1	Pituitary gland Pool	5.0
Renal ca. 786-0	0.0	Salivary Gland	0.0
Renal ca. A498	0.0	Thyroid (female)	0.0
Renal ca. ACHN	1.1	Pancreatic ca. CAPAN2	0.0
Renal ca. UO-31	0.0	Pancreas Pool	5.7

**CNS\_neurodegeneration\_v1.0 Summary:** Ag5798 This panel does not show differential expression of this gene in Alzheimer's disease. However, this profile confirms the expression of this gene at moderate levels in the brain. Please see Panel 1.5 for discussion of utility of this gene in the central nervous system.

- 5 **General\_screening\_panel\_v1.5 Summary:** Ag5798 Highest expression of this gene is seen in the thalamus (CT=31.3). This gene is also expressed at low to significant levels in the amygdala, hippocampus, cerebral cortex, substantia nigra, and whole and fetal brain samples. Therefore, therapeutic modulation of the expression or function of this gene may be useful in the treatment of neurological disorders, such as Alzheimer's disease,
- 10 Parkinson's disease, schizophrenia, multiple sclerosis, stroke and epilepsy.

In addition, this gene is expressed at much higher levels in fetal liver tissue (CT=32.5) when compared to expression in the adult counterpart (CT=37). Thus, expression of this gene may be used to differentiate between the fetal and adult source of this tissue.

- Moderate expression is also seen in a single lung cancer cell line (CT=31). Thus,  
 5 expression of this gene could be used as a marker to detect the presence of lung cancer. Furthermore, therapeutic modulation of the expression or function of this gene may be effective in the treatment of lung cancer.

**Panel 4.1D Summary:** Ag5798 Expression of this gene is low/undetectable in all samples on this panel (CTs>35). (Data not shown.)

- 10 **Panel 5 Islet Summary:** Ag5798 Expression of this gene is low/undetectable in all samples on this panel (CTs>35). (Data not shown.)

**U. CG153818-01: kinesin 19A.**

- Expression of gene CG153818-01 was assessed using the primer-probe set Ag5692,  
 described in Table UA. Results of the RTQ-PCR runs are shown in Tables UB, UC and  
 15 UD.

Table UA. Probe Name Ag5692

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5' -cgacaagggtagcaacaagtac-3'	22	1149	531
Probe	TET-5' -atcaactatcgcgacagcaagctcac-3' - TAMRA	26	1171	532
Reverse	5' -gtttcctcccagagagtcctt-3'	21	1207	533

Table UB. CNS\_neurodegeneration\_v1.0

Tissue Name	Rel. Exp.(%) Ag5692, Run 247018768	Rel. Exp.(%) Ag5692, Run 264979292	Tissue Name	Rel. Exp.(%) Ag5692, Run 247018768	Rel. Exp.(%) Ag5692, Run 264979292
AD.1 Hippo	5.1	5.3	Control (Path) 3 Temporal Ctx	5.6	6.9
AD.2 Hippo	23.3	26.6	Control (Path) 4	5.8	4.9

			Temporal Ctx		
AD 3 Hippo	4.1	5.2	AD 1 Occipital Ctx	2.9	6.2
AD 4 Hippo	19.1	22.8	AD 2 Occipital Ctx (Missing)	0.0	0.0
AD 5 Hippo	28.9	39.8	AD 3 Occipital Ctx	5.4	5.9
AD 6 Hippo	74.7	88.3	AD 4 Occipital Ctx	33.9	30.4
Control 2 Hippo	19.8	27.0	AD 5 Occipital Ctx	9.5	14.2
Control 4 Hippo	10.7	11.6	AD 6 Occipital Ctx	13.3	14.9
Control (Path) 3 Hippo	6.9	7.9	Control 1 Occipital Ctx	2.4	2.8
AD 1 Temporal Ctx	10.4	17.2	Control 2 Occipital Ctx	27.2	21.5
AD 2 Temporal Ctx	18.0	17.6	Control 3 Occipital Ctx	8.2	8.2
AD 3 Temporal Ctx	2.7	8.5	Control 4 Occipital Ctx	9.7	12.9
AD 4 Temporal Ctx	29.1	33.4	Control (Path) 1 Occipital Ctx	17.0	0.0
AD 5 Inf Temporal Ctx	100.0	100.0	Control (Path) 2 Occipital Ctx	3.7	5.8
AD 5 Sup Temporal Ctx	66.4	67.8	Control (Path) 3 Occipital Ctx	5.8	5.8
AD 6 Inf Temporal	94.6	93.3	Control (Path) 4	3.6	5.0

Ctx			Occipital Ctx		
AD 6 Sup Temporal Ctx	59.0	72.2	Control 1 Parietal Ctx	3.8	5.2
Control 1 Temporal Ctx	2.2	2.6	Control 2 Parietal Ctx	68.8	90.8
Control 2 Temporal Ctx	17.9	21.8	Control 3 Parietal Ctx	6.0	9.7
Control 3 Temporal Ctx	4.9	6.3	Control (Path) 1 Parietal Ctx	10.2	8.2
Control 3 Temporal Ctx	8.9	9.0	Control (Path) 2 Parietal Ctx	7.5	6.8
Control (Path) 1 Temporal Ctx	8.0	11.1	Control (Path) 3 Parietal Ctx	3.8	5.4
Control (Path) 2 Temporal Ctx	7.3	6.5	Control (Path) 4 Parietal Ctx	6.8	6.4

Table UC. General\_screening\_panel\_v1.5

Tissue Name	Rel. Exp.(%) Ag5692, Run 245274428	Tissue Name	Rel. Exp.(%) Ag5692, Run 245274428
Adipose	2.6	Renal ca. TK-10	0.0
Melanoma* Hs688(A).T	0.0	Bladder	14.0
Melanoma* Hs688(B).T	0.0	Gastric ca. (liver met.) NCI-N87	0.0
Melanoma* M14	0.0	Gastric ca. KATO III	0.0
Melanoma* LOXIMVI	0.0	Colon ca. SW-948	0.6
Melanoma* SK-MEL-5	0.3	Colon ca. SW480	0.4
Squamous cell carcinoma SCC-4	0.0	Colon ca.* (SW480 met) SW620	4.8
Testis Pool	7.1	Colon ca. HT29	2.0
Prostate ca.* (bone met) PC-3	0.0	Colon ca. HCT-116	0.4

Prostate Pool	0.6	Colon ca. CaCo-2	0.4
Placenta	0.0	Colon cancer tissue	1.9
Uterus Pool	1.8	Colon ca. SW1116	0.0
Ovarian ca. OVCAR-3	1.3	Colon ca. Colo-205	0.0
Ovarian ca. SK-OV-3	0.0	Colon ca. SW-48	0.7
Ovarian ca. OVCAR-4	0.0	Colon Pool	0.3
Ovarian ca. OVCAR-5	0.0	Small Intestine Pool	3.2
Ovarian ca. IGROV-1	0.7	Stomach Pool	3.6
Ovarian ca. OVCAR-8	0.0	Bone Marrow Pool	2.0
Ovary	2.3	Fetal Heart	0.3
Breast ca. MCF-7	0.0	Heart Pool	0.5
Breast ca. MDA-MB-231	0.0	Lymph Node Pool	0.9
Breast ca. BT 549	0.0	Fetal Skeletal Muscle	2.7
Breast ca. T47D	0.0	Skeletal Muscle Pool	0.7
Breast ca. MDA-N	0.0	Spleen Pool	54.7
Breast Pool	0.9	Thymus Pool	9.9
Trachea	51.1	CNS cancer (glio/astro) U87-MG	0.0
Lung	0.6	CNS cancer (glio/astro) U-118-MG	0.0
Fetal Lung	52.9	CNS cancer (neuro;met) SK-N-AS	0.0
Lung ca. NCI-N417	0.0	CNS cancer (astro) SF-539	0.0
Lung ca. LX-1	15.2	CNS cancer (astro) SNB-75	0.0
Lung ca. NCI-H146	0.0	CNS cancer (glio) SNB-19	1.0
Lung ca. SHP-77	100.0	CNS cancer (glio) SF-295	0.6
Lung ca. A549	0.0	Brain (Amygdala) Pool	27.2
Lung ca. NCI-H526	0.4	Brain (cerebellum)	8.2
Lung ca. NCI-H23	2.9	Brain (fetal)	3.1
Lung ca. NCI-H460	0.0	Brain (Hippocampus) Pool	26.2
Lung ca. HOP-62	0.0	Cerebral Cortex Pool	15.9
Lung ca. NCI-H522	0.0	Brain (Substantia	15.4

		nigra) Pool	
Liver	2.6	Brain (Thalamus) Pool	35.1
Fetal Liver	2.5	Brain (whole)	11.3
Liver ca. HepG2	0.0	Spinal Cord Pool	16.2
Kidney Pool	1.9	Adrenal Gland	1.2
Fetal Kidney	1.6	Pituitary gland Pool	0.2
Renal ca. 786-0	0.0	Salivary Gland	3.9
Renal ca. A498	0.0	Thyroid (female)	15.8
Renal ca. ACHN	0.0	Pancreatic ca. CAPAN2	0.0
Renal ca. UO-31	0.0	Pancreas Pool	4.4

Table UD. Panel 4.1D

Tissue Name	Rel. Exp.(%) Ag5692, Run 246504798	Tissue Name	Rel. Exp.(%) Ag5692, Run 246504798
Secondary Th1 act	0.0	HUVEC IL-1beta	2.0
Secondary Th2 act	1.4	HUVEC IFN gamma	100.0
Secondary Tr1 act	0.0	HUVEC TNF alpha + IFN gamma	0.0
Secondary Th1 rest	0.8	HUVEC TNF alpha + IL4	0.0
Secondary Th2 rest	0.0	HUVEC IL-11	0.0
Secondary Tr1 rest	0.0	Lung Microvascular EC none	0.0
Primary Th1 act	0.0	Lung Microvascular EC TNFalpha + IL-1beta	0.0
Primary Th2 act	0.0	Microvascular Dermal EC none	0.0
Primary Tr1 act	0.0	Microvascular Dermal EC TNFalpha + IL-1beta	0.0
Primary Th1 rest	0.0	Bronchial epithelium TNFalpha + IL1beta	0.0
Primary Th2 rest	0.0	Small airway epithelium none	0.0
Primary Tr1 rest	0.0	Small airway epithelium TNFalpha + IL-1beta	0.0
CD45RA CD4 lymphocyte act	0.0	Coronary artery SMC rest	0.0
CD45RO CD4 lymphocyte act	0.0	Coronary artery SMC TNFalpha + IL-1beta	0.0
CD8 lymphocyte act	0.0	Astrocytes rest	0.0
Secondary CD8 lymphocyte rest	0.0	Astrocytes TNFalpha + IL-1beta	0.0

Secondary CD8 lymphocyte act	0.0	KU-812 (Basophil) rest	7.0
CD4 lymphocyte none	1.4	KU-812 (Basophil) PMA/ionomycin	11.0
2ry Th1/Th2/Tr1_anti-CD95 CH11	0.0	CCD1106 (Keratinocytes) none	0.0
LAK cells rest	0.0	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	2.3
LAK cells IL-2	3.2	Liver cirrhosis	5.1
LAK cells IL-2+IL-12	0.0	NCI-H292 none	0.0
LAK cells IL-2+IFN gamma	0.0	NCI-H292 IL-4	0.0
LAK cells IL-2+ IL-18	0.0	NCI-H292 IL-9	0.0
LAK cells PMA/ionomycin	0.0	NCI-H292 IL-13	0.0
NK Cells IL-2 rest	29.7	NCI-H292 IFN gamma	0.0
Two Way MLR 3 day	0.0	HPAEC none	1.2
Two Way MLR 5 day	0.0	HPAEC TNF alpha + IL-1 beta	0.0
Two Way MLR 7 day	0.0	Lung fibroblast none	0.0
PBMC rest	2.3	Lung fibroblast TNF alpha + IL-1 beta	0.0
PBMC PWM	0.0	Lung fibroblast IL-4	0.0
PBMC PHA-L	0.0	Lung fibroblast IL-9	0.0
Ramos (B cell) none	0.0	Lung fibroblast IL-13	0.0
Ramos (B cell) ionomycin	0.0	Lung fibroblast IFN gamma	0.0
B lymphocytes PWM	0.0	Dermal fibroblast CCD1070 rest	0.0
B lymphocytes CD40L and IL-4	0.0	Dermal fibroblast CCD1070 TNF alpha	1.3
EOL-1 dbcAMP	0.0	Dermal fibroblast CCD1070 IL-1 beta	0.0
EOL-1 dbcAMP PMA/ionomycin	0.0	Dermal fibroblast IFN gamma	0.0
Dendritic cells none	0.0	Dermal fibroblast IL-4	0.0
Dendritic cells LPS	0.0	Dermal Fibroblasts rest	0.0
Dendritic cells anti-CD40	0.0	Neutrophils TNFa+LPS	0.0
Monocytes rest	0.0	Neutrophils rest	0.0
Monocytes LPS	0.0	Colon	0.0
Macrophages rest	0.0	Lung	3.6
Macrophages LPS	0.0	Thymus	1.3

HUVEC none	0.0	Kidney	0.0
HUVEC starved	0.0		

**CNS\_neurodegeneration\_v1.0 Summary:** Ag5692 Two experiments with the same probe and primer set produce results that are in excellent agreement. This panel confirms the expression of this gene at moderate levels in the brain in an independent group of individuals. This gene is found to be upregulated in the temporal cortex of Alzheimer's disease patients. This gene encodes a putative kinesin, a microtubule-based motor protein involved in the transport of organelles. Axonal transport of APP in neurons is mediated by binding with kinesin. (Gunewardena S, Neuron 2001 Nov. 8;32(3):389-401). Kamal et al. suggest that impaired APP transport leads to enhanced axonal generation and deposition of Abeta, resulting in disruption of neurotrophic signaling and neurodegeneration (Nature 10. 2001 Dec 6;414(6864):643-8). Thus, therapeutic modulation of the expression or function of this gene may be useful in the treatment of neurodegenerative disorders, and specifically may decrease neuronal death and be of use in the treatment of Alzheimer's disease.

**General\_screening\_panel\_v1.5 Summary:** Ag5692 Highest expression of this gene is seen in a lung cancer cell line (CT=29.4). Moderate levels of expression are also seen in fetal lung (CT=30) and interestingly, are much higher than expression of this gene in the adult counterpart (CT=32). Thus, expression of this gene could be used to differentiate between the adult and fetal source of this tissue. In addition, therapeutic modulation of the expression or function of this gene may be useful in the treatment of diseases that affect the lung, including lung cancer.

20 Moderate to low levels of expression are seen in all regions of the CNS examined. Please see CNS\_neurodegeneration\_v1.0 for discussion of utility of this gene in CNS disorders.

Low but significant levels of expression are also seen in pancreas, thyroid, fetal skeletal muscle, adipose and adult and fetal liver. This widespread expression among these tissues suggests that this gene product may play a role in normal neuroendocrine and metabolic function and that dysregulated expression of this gene may contribute to neuroendocrine disorders or metabolic diseases, such as obesity and diabetes.

**Panel 4.1D Summary:** Ag5692 Expression of this gene is limited to a few samples in this panel, with highest expression in IFN-gamma treated HUVEC cells (CT=31.2). Low but

significant levels of expression are also seen in PMA/ionomycin treated basophils and resting NK cells. This expression profile suggests that expression of this gene could be a marker of activated HUVEC cells. In addition, modulation of the expression or function of this gene product may reduce or eliminate the symptoms in patients with autoimmune and inflammatory diseases that involve endothelial cells, such as lupus erythematosus, asthma, emphysema, Crohn's disease, ulcerative colitis, rheumatoid arthritis, osteoarthritis, and psoriasis.

#### V. CG154435-01: Dynein beta chain, ciliary.

Expression of gene CG154435-01 was assessed using the primer-probe set Ag5694, described in Table VA. Results of the RTQ-PCR runs are shown in Tables VB, VC, VD, VE and VF.

Table VA. Probe Name Ag5694

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5' - ccaccaagtggaaagatatcaa-3'	22	3932	534
Probe	TET-5' - ccttggcaaacttcttacaatctatgtcca-3' - TAMRA	30	3965	535
Reverse	5' - ccttgtccaaagacctcatgt-3'	21	3995	536

Table VB. AI\_comprehensive panel\_v1.0

Tissue Name	Rel. Exp.(%) Ag5694, Run 245243118	Tissue Name	Rel. Exp.(%) Ag5694, Run 245243118
110967 COPD-F	0.3	112427 Match Control Psoriasis-F	0.0
110980 COPD-F	0.0	112418 Psoriasis-M	6.8
110968 COPD-M	0.2	112723 Match Control Psoriasis-M	2.6
110977 COPD-M	0.0	112419 Psoriasis-M	2.7
110989 Emphysema-F	0.1	112424 Match Control Psoriasis-M	2.9
110992 Emphysema-F	0.0	112420 Psoriasis-M	0.6
110993 Emphysema-F	0.0	112425 Match Control Psoriasis-M	2.3
110994 Emphysema-F	0.0	104689 (MF) OA Bone-Backus	0.2

110995 Emphysema-F	0.4	104690 (MF) Adj "Normal" Bone-Backus	2.6
110996 Emphysema-F	0.7	104691 (MF) OA Synovium-Backus	0.7
110997 Asthma-M	0.3	104692 (BA) OA Cartilage-Backus	2.0
111001 Asthma-F	0.0	104694 (BA) OA Bone-Backus	0.3
111002 Asthma-F	0.0	104695 (BA) Adj "Normal" Bone-Backus	0.4
111003 Atopic Asthma-F	0.0	104696 (BA) OA Synovium-Backus	0.0
111004 Atopic Asthma-F	0.1	104700 (SS) OA Bone-Backus	0.0
111005 Atopic Asthma-F	0.0	104701 (SS) Adj "Normal" Bone-Backus	1.5
111006 Atopic Asthma-F	0.0	104702 (SS) OA Synovium-Backus	2.6
111417 Allergy-M	1.0	117093 OA Cartilage Rep7	0.2
112347 Allergy-M	0.0	112672 OA Bone5	0.1
112349 Normal Lung-F	0.5	112673 OA Synovium5	2.7
112357 Normal Lung-F	0.0	112674 OA Synovial Fluid cells5	0.2
112354 Normal Lung-M	9.7	117100 OA Cartilage Rep14	3.1
112374 Crohns-F	0.0	112756 OA Bone9	1.6
112389 Match Control Crohns-F	0.2	112757 OA Synovium9	0.0
112375 Crohns-F	0.5	112758 OA Synovial Fluid Cells9	0.4
112732 Match Control Crohns-F	0.2	117125 RA Cartilage Rep2	1.5
112725 Crohns-M	0.0	113492 Bone2 RA	0.0
112387 Match Control Crohns-M	0.0	113493 Synovium2 RA	0.9
112378 Crohns-M	3.6	113494 Syn Fluid Cells RA	0.9
112390 Match Control Crohns-M	0.0	113499 Cartilage4 RA	51.4
112726 Crohns-M	0.0	113500 Bone4 RA	82.4

112731 Match Control Crohns-M	0.2	113501 Synovium4 RA	13.1
112380 Ulcer Col-F	0.0	113502 Syn Fluid Cells4 RA	0.0
112734 Match Control Ulcer Col-F	0.5	113495 Cartilage3 RA	14.3
112384 Ulcer Col-F	0.0	113496 Bone3 RA	3.1
112737 Match Control Ulcer Col-F	0.0	113497 Synovium3 RA	0.3
112386 Ulcer Col-F	100.0	113498 Syn Fluid Cells3 RA	0.6
112738 Match Control Ulcer Col-F	3.0	117106 Normal Cartilage Rep20	42.3
112381 Ulcer Col-M	0.2	113663 Bone3 Normal	0.4
112735 Match Control Ulcer Col-M	2.2	113664 Synovium3 Normal	0.4
112382 Ulcer Col-M	0.2	113665 Syn Fluid Cells3 Normal	0.2
112394 Match Control Ulcer Col-M	0.0	117107 Normal Cartilage Rep22	7.9
112383 Ulcer Col-M	0.3	113667 Bone4 Normal	0.0
112736 Match Control Ulcer Col-M	0.1	113668 Synovium4 Normal	0.0
112423 Psoriasis-F	0.4	113669 Syn Fluid Cells4 Normal	0.0

Table VC. CNS\_neurodegeneration\_v1.0

Tissue Name	Rel. Exp.(%) Ag5694, Run 247018769	Tissue Name	Rel. Exp.(%) Ag5694, Run 247018769
AD 1 Hippo	0.0	Control (Path) 3 Temporal Ctx	0.0
AD 2 Hippo	11.4	Control (Path) 4 Temporal Ctx	15.7
AD 3 Hippo	0.0	AD 1 Occipital Ctx	0.0
AD 4 Hippo	4.5	AD 2 Occipital Ctx (Missing)	0.0
AD 5 hippo	0.0	AD 3 Occipital Ctx	4.3
AD 6 Hippo	33.0	AD 4 Occipital Ctx	7.1
Control 2 Hippo	0.0	AD 5 Occipital Ctx	0.0
Control 4 Hippo	0.0	AD 6 Occipital	25.5

		Ctx	
Control (Path) 3 Hippo	0.0	Control 1 Occipital Ctx	0.0
AD 1 Temporal Ctx	7.2	Control 2 Occipital Ctx	30.6
AD 2 Temporal Ctx	17.3	Control 3 Occipital Ctx	6.4
AD 3 Temporal Ctx	7.1	Control 4 Occipital Ctx	5.1
AD 4 Temporal Ctx	0.0	Control (Path) 1 Occipital Ctx	6.4
AD 5 Inf Temporal Ctx	7.4	Control (Path) 2 Occipital Ctx	0.0
AD 5 Sup Temporal Ctx	6.4	Control (Path) 3 Occipital Ctx	0.0
AD 6 Inf Temporal Ctx	19.6	Control (Path) 4 Occipital Ctx	13.1
AD 6 Sup Temporal Ctx	100.0	Control 1 Parietal Ctx	0.0
Control 1 Temporal Ctx	0.0	Control 2 Parietal Ctx	5.0
Control 2 Temporal Ctx	0.0	Control 3 Parietal Ctx	5.7
Control 3 Temporal Ctx	0.0	Control (Path) 1 Parietal Ctx	7.7
Control 4 Temporal Ctx	21.0	Control (Path) 2 Parietal Ctx	13.6
Control (Path) 1 Temporal Ctx	6.4	Control (Path) 3 Parietal Ctx	4.1
Control (Path) 2 Temporal Ctx	13.0	Control (Path) 4 Parietal Ctx	2.1

Table VD. General\_screening\_panel\_v1.5

Tissue Name	Rel. Exp.(%) Ag5694, Run 249040574	Tissue Name	Rel. Exp.(%) Ag5694, Run 249040574
Adipose	0.6	Renal ca. TK-10	40.3
Melanoma* Hs688(A).T	0.0	Bladder	1.4
Melanoma* Hs688(B).T	0.0	Gastric ca. (liver met.) NCI-N87	2.8
Melanoma* M14	0.0	Gastric ca. KATO III	1.1
Melanoma* LOXIMVI	0.9	Colon ca. SW-948	0.0

Melanoma* SK-MEL-5	11.0	Colon ca. SW480	1.6
Squamous cell carcinoma SCC-4	2.3	Colon ca.* (SW480 met) SW620	0.6
Testis Pool	100.0	Colon ca. HT29	0.0
Prostate ca.* (bone met) PC-3	0.0	Colon ca. HCT-116	2.8
Prostate Pool	0.0	Colon ca. CaCo-2	0.0
Placenta	4.3	Colon cancer tissue	2.5
Uterus Pool	0.4	Colon ca. SW1116	0.0
Ovarian ca. OVCAR-3	5.2	Colon ca. Colo-205	0.0
Ovarian ca. SK-OV-3	3.3	Colon ca. SW-48	0.0
Ovarian ca. OVCAR-4	2.3	Colon Pool	0.0
Ovarian ca. OVCAR-5	1.4	Small Intestine Pool	0.0
Ovarian ca. IGROV-1	1.2	Stomach Pool	0.0
Ovarian ca. OVCAR-8	1.6	Bone Marrow Pool	0.0
Ovary	0.0	Fetal Heart	0.0
Breast ca. MCF-7	0.9	Heart Pool	0.0
Breast ca. MDA-MB-231	0.5	Lymph Node Pool	0.0
Breast ca. BT 549	0.0	Fetal Skeletal Muscle	0.0
Breast ca. T47D	0.0	Skeletal Muscle Pool	0.3
Breast ca. MDA-N	0.0	Spleen Pool	0.5
Breast Pool	0.8	Thymus Pool	3.0
Trachea	2.6	CNS cancer (glio/astro) U87-MG	0.9
Lung	0.0	CNS cancer (glio/astro) U-118-MG	0.3
Fetal Lung	12.7	CNS cancer (neuro;met) SK-N-AS	0.9
Lung ca. NCI-N417	0.0	CNS cancer (astro) SF-539	0.0
Lung ca. LX-1	13.5	CNS cancer (astro) SNB-75	0.3
Lung ca. NCI-H146	0.5	CNS cancer (glio) SNB-19	0.9
Lung ca. SHP-77	8.6	CNS cancer (glio) SF-295	0.0

Lung ca. A549	1.2	Brain (Amygdala) Pool	0.3
Lung ca. NCI-H526	0.0	Brain (cerebellum)	0.5
Lung ca. NCI-H23	41.8	Brain (fetal)	0.0
Lung ca. NCI-H460	0.6	Brain (Hippocampus) Pool	0.2
Lung ca. HOP-62	0.6	Cerebral Cortex Pool	0.9
Lung ca. NCI-H522	0.0	Brain (Substantia nigra) Pool	0.8
Liver	0.0	Brain (Thalamus) Pool	0.9
Fetal Liver	0.5	Brain (whole)	0.0
Liver ca. HepG2	90.1	Spinal Cord Pool	0.3
Kidney Pool	0.0	Adrenal Gland	0.2
Fetal Kidney	0.9	Pituitary gland Pool	0.0
Renal ca. 786-0	0.6	Salivary Gland	0.2
Renal ca. A498	1.0	Thyroid (female)	0.6
Renal ca. ACHN	0.7	Pancreatic ca. CAPAN2	6.7
Renal ca. UO-31	1.3	Pancreas Pool	0.0

Table VE. Panel 4.1D

Tissue Name	Rel. Exp.(%) Ag5694, Run 246504805	Tissue Name	Rel. Exp.(%) Ag5694, Run 246504805
Secondary Th1 act	0.0	HUVEC IL-1beta	0.5
Secondary Th2 act	0.0	HUVEC IFN gamma	0.0
Secondary Tr1 act	0.0	HUVEC TNF.alpha + IFN gamma	0.0
Secondary Th1 rest	0.0	HUVEC TNF.alpha + IL4	0.0
Secondary Th2 rest	0.0	HUVEC IL-11	0.0
Secondary Tr1 rest	0.0	Lung Microvascular EC none	0.0
Primary Th1 act	0.0	Lung Microvascular EC TNF.alpha + IL-1beta	0.0
Primary Th2 act	0.0	Microvascular Dermal EC none	0.0
Primary Tr1 act	0.0	Microvascular Dermal EC TNF.alpha + IL-1beta	0.0
Primary Th1 rest	0.0	Bronchial epithelium TNF.alpha + IL1beta	0.0
Primary Th2 rest	0.0	Small airway epithelium none	0.0
Primary Tr1 rest	0.0	Small airway epithelium TNF.alpha + IL-1beta	0.0

CD45RA CD4 lymphocyte act	0.0	Coronary artery. SMC rest	0.0
CD45RO CD4 lymphocyte act	0.4	Coronary artery. SMC TNFalpha + IL-1beta	0.0
CD8 lymphocyte act	0.0	Astrocytes rest	0.0
Secondary CD8 lymphocyte rest	0.9	Astrocytes TNFalpha + IL-1beta	0.0
Secondary CD8 lymphocyte act	0.0	KU-812 (Basophil) rest	0.0
CD4 lymphocyte none	0.0	KU-812 (Basophil) PMA/ionomycin	0.0
2ry Th1/Th2/Tr1_anti-CD95 CH1 1	0.0	CCD1106 (Keratinocytes) none	1.1
LAK cells rest	0.4	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	1.8
LAK cells IL-2	0.0	Liver cirrhosis	0.6
LAK cells IL-2+IL-12	0.0	NCI-H292 none	2.1
LAK cells IL-2+IFN gamma	0.0	NCI-H292 IL-4	1.4
LAK cells IL-2+ IL-18	0.0	NCI-H292 IL-9	0.0
LAK cells PMA/ionomycin	5.8	NCI-H292 IL-13	0.0
NK Cells IL-2 rest	0.4	NCI-H292 IFN gamma	0.4
Two Way MLR 3 day	0.0	HPAEC none	0.0
Two Way MLR 5 day	0.0	HPAEC TNF alpha + IL-1 beta	0.0
Two Way MLR 7 day	0.0	Lung fibroblast none	0.0
PBMC rest	0.0	Lung fibroblast TNF alpha + IL-1 beta	0.0
PBMC PWM	0.0	Lung fibroblast IL-4	0.0
PBMC PHA-L	0.0	Lung fibroblast IL-9	0.0
Ramos (B cell) none	0.0	Lung fibroblast IL-13	0.0
Ramos (B cell) ionomycin	0.0	Lung fibroblast IFN gamma	0.0
B lymphocytes PWM	0.0	Dermal fibroblast CCD1070 rest	0.0
B lymphocytes CD40L and IL-4	0.4	Dermal fibroblast CCD1070 TNF alpha	0.0
EOL-1 dbcAMP	0.0	Dermal fibroblast CCD1070 IL-1 beta	0.0
EOL-1 dbcAMP PMA/ionomycin	0.9	Dermal fibroblast IFN gamma	0.0
Dendritic cells none	0.0	Dermal fibroblast IL-4	0.0

Dendritic cells LPS	0.0	Dermal Fibroblasts rest	0.0
Dendritic cells anti-CD40	0.0	Neutrophils TNFa+LPS	13.1
Monocytes rest	0.0	Neutrophils rest	0.4
Monocytes LPS	100.0	Colon	0.0
Macrophages rest	0.0	Lung	0.0
Macrophages LPS	0.6	Thymus	0.0
HUVEC none	0.0	Kidney	2.4
HUVEC starved	0.0		

Table VF. Panel 5 Islet

Tissue Name	Rel. Exp.(%) Ag5694, Run 253330720	Tissue Name	Rel. Exp.(%) Ag5694, Run 253330720
97457_Patient-02go_adipose	0.0	94709_Donor 2 AM - A_adipose	0.0
97476_Patient-07sk_skeletal muscle	0.0	94710_Donor 2 AM - B_adipose	0.0
97477_Patient-07ut_uterus	0.0	94711_Donor 2 AM - C_adipose	0.0
97478_Patient-07pl_placenta	0.0	94712_Donor 2 AD - A_adipose	0.0
99167_Bayer Patient 1	67.8	94713_Donor 2 AD - B_adipose	0.0
97482_Patient-08ut_uterus	0.0	94714_Donor 2 AD - C_adipose	0.0
97483_Patient-08pl_placenta	12.2	94742_Donor 3 U - A_Mesenchymal Stem Cells	0.0
97486_Patient-09sk_skeletal muscle	5.5	94743_Donor 3 U - B_Mesenchymal Stem Cells	0.0
97487_Patient-09ut_uterus	0.0	94730_Donor 3 AM - A_adipose	0.0
97488_Patient-09pl_placenta	7.4	94731_Donor 3 AM - B_adipose	0.0
97492_Patient-10ut_uterus	0.0	94732_Donor 3 AM - C_adipose	0.0
97493_Patient-10pl_placenta	0.0	94733_Donor 3 AD - A_adipose	0.0
97495_Patient-11go_adipose	0.0	94734_Donor 3 AD - B_adipose	7.6
97496_Patient-11sk_skeletal muscle	0.0	94735_Donor 3 AD - C_adipose	0.0
97497_Patient-	0.0	77138_Liver_HepG2untreated	100.0

11ut_uterus			
97498_Patient-11pl_placenta	16.7	73556_Heart_Cardiac stromal cells (primary)	0.0
97500_Patient-12go_adipose	0.0	81735_Small Intestine	0.0
97501_Patient-12sk_skeletal muscle	0.0	72409_Kidney_Proximal Convoluted Tubule	0.0
97502_Patient-12ut_uterus	0.0	82685_Small intestine_Duodenum	0.0
97503_Patient-12pl_placenta	8.4	90650_Adrenal_Adrenocortical adenoma	0.0
94721_Donor 2 U - A_Mesenchymal Stem Cells	0.0	72410_Kidney_HRCE	7.6
94722_Donor 2 U - B_Mesenchymal Stem Cells	0.0	72411_Kidney_HRE	0.0
94723_Donor 2 U - C_Mesenchymal Stem Cells	0.0	73139_Uterus_Uterine smooth muscle cells	0.0

- AI\_comprehensive\_panel\_v1.0 Summary:** Ag5694 Highest expression of this gene is detected in ulcerative colitis sample (CT=30.2). Interestingly, expression of this gene is higher in ulcerative colitis sample as compared to matching control sample (CT=35). Therefore, expression of this gene may be used to distinguish between these two samples and also as a marker to detect ulcerative colitis. In addition, moderate expression of this gene is seen in cartilage, bone and synovium from rheumatoid arthritis patient, low expression in normal lung, psoriasis, and normal cartilage Rep22. Therefore, therapeutic modulation of this gene may be useful in the treatment of rheumatoid arthritis, ulcerative colitis, and psoriasis.
- 5
- 10 **CNS\_neurodegeneration\_v1.0 Summary:** Ag5694 Low expression of this gene is detected in temporal cortex of an Alzheimer's patient. Therefore, therapeutic modulation of this gene may be useful in the treatment of Alzheimer's disease.

- General\_screening\_panel\_v1.5 Summary:** Ag5694 Highest expression of this gene is detected in testis (CT=29.8). Therefore, expression of this gene may be used to differentiate testis from other samples in this panel. In addition, therapeutic modulation of this gene may be useful in the treatment of testis related diseases including fertility and hypogonadism.
- 15

In addition, moderate to low levels of expression of this gene is detected in number of cancer cell lines derived from melanoma, pancreatic, renal, liver, lung, and ovarian cancers. Therefore, expression of this gene may be used as diagnostic marker to detect these cancers and also, therapeutic modulation of this gene through the use of antibodies or small molecule drug may be useful in the treatment of melanoma, pancreatic, renal, liver, lung, and ovarian cancers.

**Panel 4.1D Summary:** Ag5694 Moderate expression of this gene is detected mainly in LPS treated monocytes (CT=29.9). In addition, low levels of expression of this gene is also seen in TNF alpha and LPS treated neutrophils. Therefore, expression of this gene may be used to distinguish activated monocytes and neutrophils from other samples in this panel. The expression of this gene in LPS treated monocytes, cells that play a crucial role in linking innate immunity to adaptive immunity, suggests a role for this gene product in initiating inflammatory reactions. Therefore, modulation of the expression or activity of this gene through the application of monoclonal antibodies may reduce or prevent early stages of inflammation and reduce the severity of inflammatory diseases such as psoriasis, asthma, inflammatory bowel disease, rheumatoid arthritis, osteoarthritis and other lung inflammatory diseases. In addition, small molecule or antibody antagonists of this gene product may be effective in increasing the immune response in patients with AIDS or other immunodeficiencies.

**Panel 5 Islet Summary:** Ag5694 Low levels of expression of this gene is exclusively seen in liver cancer HepG2 cell line (CT=34.7). Please see panel 1.5 for further utility of this gene.

**W. CG154465-01: kinesin 18B.**

Expression of gene CG154465-01 was assessed using the primer-probe set Ag5695, described in Table WA. Results of the RTQ-PCR runs are shown in Tables WB and WC.

Table WA. Probe Name Ag5695

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5' - tcaatgccacctttgatctct - 3'	21	2279	537
Probe	TET-5' - aaagcccagtttccatgaatgcattg - 3' - TAMRA	26	2316	538

Reverse	5'-cagctcctgggggtattttgt-3'	20	2348	539
---------	-----------------------------	----	------	-----

Table WB. General\_screening\_panel\_v1.5

Tissue Name	Rel. Exp.(%) Ag5695, Run 245274429	Tissue Name	Rel. Exp.(%) Ag5695, Run 245274429
Adipose	0.1	Renal ca. TK-10	24.0
Melanoma* Hs688(A).T	0.5	Bladder	3.1
Melanoma* Hs688(B).T	1.2	Gastric ca. (liver met.) NCI-N87	5.4
Melanoma* M14	43.2	Gastric ca. KATO III	97.9
Melanoma* LOXIMVI	45.7	Colon ca. SW-948	24.8
Melanoma* SK- MEL-5	17.3	Colon ca. SW480	86.5
Squamous cell carcinoma SCC-4	14.6	Colon ca.* (SW480 met) SW620	37.6
Testis Pool	1.0	Colon ca. HT29	17.4
Prostate ca.* (bone met) PC-3	2.2	Colon ca. HCT-116	100.0
Prostate Pool	0.3	Colon ca. CaCo-2	31.4
Placenta	1.5	Colon cancer tissue	7.0
Uterus Pool	0.3	Colon ca. SW1116	16.8
Ovarian ca. OVCAR-3	39.5	Colon ca. Colo-205	18.2
Ovarian ca. SK- OV-3	82.4	Colon ca. SW-48	11.0
Ovarian ca. OVCAR-4	23.7	Colon Pool	0.6
Ovarian ca. OVCAR-5	33.0	Small Intestine Pool	0.2
Ovarian ca. IGROV-1	9.3	Stomach Pool	0.2
Ovarian ca. OVCAR-8	10.5	Bone Marrow Pool	0.2
Ovary	0.0	Fetal Heart	6.0
Breast ca. MCF-7	20.9	Heart Pool	0.3
Breast ca. MDA- MB-231	69.7	Lymph Node Pool	0.6
Breast ca. BT 549	50.0	Fetal Skeletal Muscle	3.0
Breast ca. T47D	24.1	Skeletal Muscle Pool	0.0
Breast ca. MDA-N	24.3	Spleen Pool	1.4

Breast Pool	0.6	Thymus Pool	12.1
Trachea	0.6	CNS cancer (glio/astro) U87-MG	19.1
Lung	0.1	CNS cancer (glio/astro) U-118-MG	97.9
Fetal Lung	7.2	CNS cancer (neuro;met) SK-N-AS	52.5
Lung ca. NCI-N417	13.9	CNS cancer (astro) SF-539	25.7
Lung ca. LX-1	25.3	CNS cancer (astro) SNB-75	66.0
Lung ca. NCI-H146	14.5	CNS cancer (glio) SNB-19	9.4
Lung ca. SHP-77	25.5	CNS cancer (glio) SF-295	5.3
Lung ca. A549	55.9	Brain (Amygdala) Pool	0.1
Lung ca. NCI-H526	14.9	Brain (cerebellum)	0.1
Lung ca. NCI-H23	22.4	Brain (fetal)	2.7
Lung ca. NCI-H460	0.0	Brain (Hippocampus) Pool	0.2
Lung ca. HOP-62	5.4	Cerebral Cortex Pool	0.3
Lung ca. NCI-H522	34.6	Brain (Substantia nigra) Pool	0.1
Liver	0.0	Brain (Thalamus) Pool	0.2
Fetal Liver	33.2	Brain (whole)	0.3
Liver ca. HepG2	12.8	Spinal Cord Pool	0.1
Kidney Pool	0.1	Adrenal Gland	0.1
Fetal Kidney	12.2	Pituitary gland Pool	0.1
Renal ca. 786-0	30.6	Salivary Gland	0.3
Renal ca. A498	4.9	Thyroid (female)	0.0
Renal ca. ACHN	12.9	Pancreatic ca. CAPAN2	41.8
Renal ca. UO-31	17.3	Pancreas Pool	0.5

Table WC. Panel 4.1D

Tissue Name	Rel. Exp.(%) Ag5695, Run 246504814	Tissue Name	Rel. Exp.(%) Ag5695, Run 246504814
Secondary Th1 act	79.6	HUVEC IL-1beta	42.9
Secondary Th2 act	74.2	HUVEC IFN gamma	27.7
Secondary Tr1 act	18.9	HUVEC TNF alpha + IFN gamma	5.7
Secondary Th1 rest	0.2	HUVEC TNF alpha + IL4	4.5

Secondary Th2 rest	0.3	HUVEC IL-11	23.2
Secondary Tr1 rest	0.0	Lung Microvascular EC none	24.8
Primary Th1 act	0.9	Lung Microvascular EC TNFalpha + IL-1beta	1.9
Primary Th2 act	38.4	Microvascular Dermal EC none	2.4
Primary Tr1 act	30.8	Microvascular Dermal EC TNFalpha + IL-1beta	4.4
Primary Th1 rest	2.0	Bronchial epithelium TNFalpha + IL1beta	1.8
Primary Th2 rest	4.2	Small airway epithelium none	1.2
Primary Tr1 rest	2.7	Small airway epithelium TNFalpha + IL-1beta	4.5
CD45RA CD4 lymphocyte act	52.5	Coronary artery SMC rest	4.4
CD45RO CD4 lymphocyte act	47.0	Coronary artery SMC TNFalpha + IL-1beta	3.2
CD8 lymphocyte act	11.4	Astrocytes rest	0.3
Secondary CD8 lymphocyte rest	24.1	Astrocytes TNFalpha + IL-1beta	0.7
Secondary CD8 lymphocyte act	4.4	KU-812 (Basophil) rest	32.1
CD4 lymphocyte none	0.0	KU-812 (Basophil) PMA/ionomycin	42.3
2ry Th1/Th2/Tr1_anti-CD95 CH11	3.5	CCD1106 (Keratinocytes) none	44.8
LAK cells rest	1.6	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	9.1
LAK cells IL-2	8.7	Liver cirrhosis	2.7
LAK cells IL-2+IL-12	1.9	NCI-H292 none	19.9
LAK cells IL-2+IFN gamma	10.5	NCI-H292 IL-4	42.9
LAK cells IL-2+ IL-18	6.3	NCI-H292 IL-9	58.6
LAK cells PMA/ionomycin	3.1	NCI-H292 IL-13	52.5
NK Cells IL-2 rest	81.2	NCI-H292 IFN gamma	20.3
Two Way MLR 3 day	1.9	HPAEC none	7.4
Two Way MLR 5 day	2.9	HPAEC TNF alpha + IL-1 beta	21.3
Two Way MLR 7 day	9.2	Lung fibroblast none	5.9
PBMC rest	0.0	Lung fibroblast TNF	8.9

		alpha + IL-1 beta	
PBMC PWM	4.0	Lung fibroblast IL-4	0.8
PBMC PHA-L	12.5	Lung fibroblast IL-9	5.8
Ramos (B cell) none	8.1	Lung fibroblast IL-13	0.4
Ramos (B cell) ionomycin	76.3	Lung fibroblast IFN gamma	1.4
B lymphocytes PWM	52.9	Dermal fibroblast CCD1070 rest	100.0
B lymphocytes CD40L and IL-4	49.7	Dermal fibroblast CCD1070 TNF alpha	93.3
EOL-1 dbcAMP	31.6	Dermal fibroblast CCD1070 IL-1 beta	40.3
EOL-1 dbcAMP PMA/ionomycin	1.9	Dermal fibroblast IFN gamma	27.9
Dendritic cells none	0.6	Dermal fibroblast IL-4	40.3
Dendritic cells LPS	0.0	Dermal Fibroblasts rest	18.3
Dendritic cells anti-CD40	0.0	Neutrophils TNFa+LPS	0.0
Monocytes rest	0.0	Neutrophils rest	0.0
Monocytes LPS	0.0	Colon	0.0
Macrophages rest	2.2	Lung	0.2
Macrophages LPS	0.2	Thymus	8.5
HUVEC none	31.0	Kidney	0.0
HUVEC starved	55.5		

**CNS\_neurodegeneration\_v1.0 Summary:** Ag5695 Expression of this gene is low/undetectable (CTs > 35) across all of the samples on this panel (data not shown).

- General\_screening\_panel\_v1.5 Summary:** Ag5695 Highest expression of this gene is detected in a colon cancer HCT-116 cell line (CT=27). Moderate expression of this gene is also seen in cluster of cancer cell lines derived from pancreatic, gastric, colon, lung, liver, renal, breast, ovarian, prostate, squamous cell carcinoma, melanoma and brain cancers. Thus, expression of this gene could be used as a marker to detect the presence of these cancers. Furthermore, therapeutic modulation of the expression or function of this gene may be effective in the treatment of pancreatic, gastric, colon, lung, liver, renal, breast, ovarian, prostate, squamous cell carcinoma, melanoma and brain cancers.

In addition, significant expression of this gene is seen in fetal tissues, including fetal lung, liver, kidney, heart, and skeletal muscle. Expression of this gene is higher in fetal (CTs=28-32) compared to corresponding adult lung, liver, kidney, heart, and skeletal muscle tissues.

Therefore, expression of this gene may be useful in distinguishing between fetal and adult lung, liver, kidney, heart, and skeletal muscle. In addition, expression in fetal tissue suggests a role for the protein encoded by this gene in growth and development of these tissues in the fetus and thus may also act in a regenerative capacity in the adult.

- 5 **Panel 4.1D Summary:** Ag5695 Highest expression of this gene is detected in dermal fibroblast (CT=29.2). Moderate to low levels of expression of this gene is detected in polarized T cells (primary and secondary Th1, Th2, and Tr1), activated CD45RA CD4 and CD45RO CD4 lymphocytes, LAK cells, resting IL-2 treated NK cells, activated PBMC cells, Ramos B cells, B lymphocytes, eosinophils, endothelial cells, basophils, NCI-H292
- 10 cells, lung and dermal fibroblasts and thymus. Interestingly, expression of this gene is upregulated in activated polarized T cells, stimulated PBMC cells, and activated Ramos B cells. Therefore, therapeutic modulation of this gene may be useful in the treatment of autoimmune and inflammatory disorders including psoriasis, allergy, asthma, inflammatory bowel disease, rheumatoid arthritis and osteoarthritis.

15 **X. CG154492-01: HIGH-AFFINITY CGMP-SPECIFIC 3',5'-CYCLIC PHOSPHODIESTERASE 9A.**

Expression of gene CG154492-01 was assessed using the primer-probe set Ag6818, described in Table XA. Results of the RTQ-PCR runs are shown in Table XB.

Table XA. Probe Name Ag6818

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5' -gcagaaattatggattctttcaaag-3'	25	1345	540
Probe	TET-5' -tcctcggtgctgtagtcaaaattctcca-3' - TAMRA	28	1376	541
Reverse	5' -ggtcgctgagggatcatg-3'	17	1407	542

20 Table XB. General\_screening\_panel\_v1.6

Tissue Name	Rel. Exp.(%) Ag6818, Run 278391557	Tissue Name	Rel. Exp.(%) Ag6818, Run 278391557
Adipose	18.4	Renal ca. TK-10	15.4
Melanoma* Hs688(A).T	0.0	Bladder	0.0
Melanoma*	0.0	Gastric ca. (liver met.)	15.4

Hs688(B).T.		NCI-N87	
Melanoma* M14	0.0	Gastric ca. KATO III	0.0
Melanoma* LOXIMVI	0.0	Colon ca. SW-948	0.0
Melanoma* SK-MEL-5	10.5	Colon ca. SW480	19.6
Squamous cell carcinoma SCC-4	0.0	Colon ca.* (SW480 met) SW620	3.2
Testis Pool	8.1	Colon ca. HT29	3.6
Prostate ca.* (bone met) PC-3	0.0	Colon ca. HCT-116	29.7
Prostate Pool	24.3	Colon ca. CaCo-2	7.9
Placenta	3.7	Colon cancer tissue	2.5
Uterus Pool	0.0	Colon ca. SW1116	0.0
Ovarian ca. OVCAR-3	53.6	Colon ca. Colo-205	0.0
Ovarian ca. SK-OV-3	31.6	Colon ca. SW-48	0.0
Ovarian ca. OVCAR-4	9.4	Colon Pool	3.0
Ovarian ca. OVCAR-5	24.7	Small Intestine Pool	5.9
Ovarian ca. IGROV-1	14.1	Stomach Pool	6.0
Ovarian ca. OVCAR-8	4.3	Bone Marrow Pool	0.0
Ovary	6.1	Fetal Heart	9.8
Breast ca. MCF-7	3.0	Heart Pool	3.1
Breast ca. MDA-MB-231	0.0	Lymph Node Pool	3.5
Breast ca. BT 549	0.0	Fetal Skeletal Muscle	3.3
Breast ca. T47D	0.0	Skeletal Muscle Pool	0.0
Breast ca. MDA-N	5.6	Spleen Pool	2.8
Breast Pool	2.2	Thymus Pool	6.0
Trachea	2.8	CNS cancer (glio/astro) U87-MG	0.0
Lung	0.0	CNS cancer (glio/astro) U-118-MG	0.0
Fetal Lung	33.2	CNS cancer (neuro;met) SK-N-AS	0.0
Lung ca. NCI-N417	0.0	CNS cancer (astro) SF-539	0.0
Lung ca. LX-1	3.1	CNS cancer (astro) SNB-75	0.0

Lung ca. NCI-H146	3.5	CNS cancer (glio) SNB-19	24.0
Lung ca. SHP-77	0.0	CNS cancer (glio) SF- 295	0.0
Lung ca. A549	0.0	Brain (Amygdala) Pool	3.7
Lung ca. NCI-H526	0.0	Brain (cerebellum)	50.7
Lung ca. NCI-H23	25.0	Brain (fetal)	100.0
Lung ca. NCI-H460	7.3	Brain (Hippocampus) Pool	2.2
Lung ca. HOP-62	0.0	Cerebral Cortex Pool	7.6
Lung ca. NCI-H522	65.1	Brain (Substantia nigra) Pool	11.8
Liver	0.0	Brain (Thalamus) Pool	15.3
Fetal Liver	4.4	Brain (whole)	20.2
Liver ca. HepG2	31.9	Spinal Cord Pool	10.1
Kidney Pool	27.2	Adrenal Gland	16.6
Fetal Kidney	10.3	Pituitary gland Pool	0.0
Renal ca. 786-0	0.0	Salivary Gland	4.3
Renal ca. A498	0.0	Thyroid (female)	0.0
Renal ca. ACHN	0.0	Pancreatic ca. CAPAN2	30.1
Renal ca. UO-31	0.0	Pancreas Pool	14.0

**CNS\_neurodegeneration\_v1.0 Summary:** Ag6818 Expression of this gene is low/undetectable in all samples on this panel (CTs>35). (Data not shown.)

**General\_screening\_panel\_v1.6 Summary:** Ag6818 Expression of this gene is limited to the fetal brain (CT=34.5). Thus, expression of this gene could be used to differentiate  
5 between fetal and adult brain tissue and as a marker of fetal neural tissue.

**Panel 4.1D Summary:** Ag6818 Expression of this gene is low/undetectable in all samples on this panel (CTs>35). (Data not shown.)

**Panel 5 Islet Summary:** Ag6818 Expression of this gene is low/undetectable in all samples on this panel (CTs>35). (Data not shown.)

#### 10 **Y. CG154509-01: CYTOPLASMIC DYNEIN HEAVY CHAIN.**

Expression of gene CG154509-01 was assessed using the primer-probe set Ag5696, described in Table YA. Results of the RTQ-PCR runs are shown in Tables YB, YC and YD.

Table YA. Probe Name Ag5696

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5' - ccagattgaagtgatgaaagga - 3'	22	3156	543
Probe	TET-5' - cacgtcttcagatctattatcaagaactgg - 3' - TAMRA	30	3188	544
Reverse	5' - gtcccaacgagctttaatttt - 3'	22	3219	545

Table YB. AI\_comprehensive panel\_v1.0

Tissue Name	Rel. Exp.(%) Ag5696, Run 245243119	Tissue Name	Rel. Exp.(%) Ag5696, Run 245243119
110967 COPD-F	20.3	112427 Match Control Psoriasis-F	21.0
110980 COPD-F	5.1	112418 Psoriasis-M	22.5
110968 COPD-M	21.3	112723 Match Control Psoriasis-M	61.1
110977 COPD-M	24.7	112419 Psoriasis-M	2.8
110989 Emphysema-F	8.3	112424 Match Control Psoriasis-M	24.7
110992 Emphysema-F	16.5	112420 Psoriasis-M	12.3
110993 Emphysema-F	18.2	112425 Match Control Psoriasis-M	25.9
110994 Emphysema-F	8.6	104689 (MF) OA Bone-Backus	29.5
110995 Emphysema-F	15.2	104690 (MF) Adj "Normal" Bone-Backus	0.6
110996 Emphysema-F	8.5	104691 (MF) OA Synovium-Backus	94.6
110997 Asthma-M	18.2	104692 (BA) OA Cartilage-Backus	21.0
111001 Asthma-F	4.5	104694 (BA) OA Bone-Backus	15.1
111002 Asthma-F	54.0	104695 (BA) Adj "Normal" Bone-Backus	31.6
111003 Atopic Asthma-F	20.6	104696 (BA) OA Synovium-Backus	11.4
111004 Atopic Asthma-F	0.0	104700 (SS) OA Bone-Backus	10.5
111005 Atopic	17.2	104701 (SS) Adj	100.0

Asthma-F		"Normal" Bone-Backus	
111006 Atopic Asthma-F	76.8	104702 (SS) OA Synovium-Backus	10.8
111417 Allergy-M	85.3	117093 OA Cartilage Rep7	9.2
112347 Allergy-M	0.0	112672 OA Bone5	4.9
112349 Normal Lung-F	5.1	112673 OA Synovium5	2.4
112357 Normal Lung-F	13.4	112674 OA Synovial Fluid cells5	12.4
112354 Normal Lung-M	89.5	117100 OA Cartilage Rep14	72.7
112374 Crohns-F	52.1	112756 OA Bone9	5.7
112389 Match Control Crohns-F	47.6	112757 OA Synovium9	0.9
112375 Crohns-F	6.2	112758 OA Synovial Fluid Cells9	21.5
112732 Match Control Crohns-F	17.7	117125 RA Cartilage Rep2	5.5
112725 Crohns-M	42.3	113492 Bone2 RA	0.0
112387 Match Control Crohns-M	18.6	113493 Synovium2 RA	10.1
112378 Crohns-M	0.3	113494 Syn Fluid Cells RA	8.9
112390 Match Control Crohns-M	19.2	113499 Cartilage4 RA	18.8
112726 Crohns-M	0.6	113500 Bone4 RA	0.5
112731 Match Control Crohns-M	4.7	113501 Synovium4 RA	5.0
112380 Ulcer Col-F	48.3	113502 Syn Fluid Cells4 RA	4.8
112734 Match Control Ulcer Col-F	9.1	113495 Cartilage3 RA	33.4
112384 Ulcer Col-F	13.2	113496 Bone3 RA	18.9
112737 Match Control Ulcer Col-F	23.5	113497 Synovium3 RA	3.9
112386 Ulcer Col-F	24.1	113498 Syn Fluid Cells3 RA	0.0
112738 Match Control Ulcer Col-F	26.4	117106 Normal Cartilage Rep20	41.2
112381 Ulcer Col-M	5.6	113663 Bone3 Normal	31.6
112735 Match Control Ulcer Col-M	14.5	113664 Synovium3 Normal	18.3

112382 Ulcer Col-M	37.1	113665 Syn Fluid Cells3 Normal	80.1
112394 Match Control Ulcer Col-M	7.1	117107 Normal Cartilage Rep22	13.3
112383 Ulcer Col-M	21.9	113667 Bone4 Normal	23.8
112736 Match Control Ulcer Col-M	44.1	113668 Synovium4 Normal	22.1
112423 Psoriasis-F	34.2	113669 Syn Fluid Cells4 Normal	20.3

Table YC. CNS\_neurodegeneration\_v1.0

Tissue Name	Rel. Exp.(%) Ag5696, Run 247018771	Rel. Exp.(%) Ag5696, Run 312325348	Tissue Name	Rel. Exp.(%) Ag5696, Run 247018771	Rel. Exp.(%) Ag5696, Run 312325348
AD 1 Hippo	9.7	45.4	Control (Path) 3 Temporal Ctx	16.2	56.6
AD 2 Hippo	33.0	93.3	Control (Path) 4 Temporal Ctx	76.8	27.7
AD 3 Hippo	17.1	43.2	AD 1 Occipital Ctx	48.6	49.3
AD 4 Hippo	24.5	42.0	AD 2 Occipital Ctx (Missing)	0.0	78.5
AD 5 hippo	100.0	33.7	AD 3 Occipital Ctx	20.9	33.9
AD 6 Hippo	45.4	100.0	AD 4 Occipital Ctx	48.3	50.3
Control 2 Hippo	34.4	62.9	AD 5 Occipital Ctx	32.1	25.0
Control 4 Hippo	27.5	26.2	AD 6 Occipital Ctx	46.7	43.2
Control (Path) 3 Hippo	24.8	25.9	Control 1 Occipital Ctx	14.3	45.4
AD 1 Temporal	42.9	28.1	Control 2	43.8	37.1

Ctx			Occipital Ctx		
AD 2 Temporal Ctx	47.6	55.9	Control 3 Occipital Ctx	57.8	31.6
AD 3 Temporal Ctx	23.5	48.3	Control 4 Occipital Ctx	20.3	39.5
AD 4 Temporal Ctx	48.6	76.3	Control (Path) 1 Occipital Ctx	99.3	22.2
AD 5 Inf Temporal Ctx	78.5	87.1	Control (Path) 2 Occipital Ctx	31.6	51.8
AD 5 Sup Temporal Ctx	50.0	45.7	Control (Path) 3 Occipital Ctx	5.1	60.3
AD 6 Inf Temporal Ctx	50.3	47.6	Control (Path) 4 Occipital Ctx	69.7	20.9
AD 6 Sup Temporal Ctx	86.5	13.9	Control 1 Parietal Ctx	23.3	29.9
Control 1 Temporal Ctx	21.6	21.2	Control 2 Parietal Ctx	56.3	37.4
Control 2 Temporal Ctx	29.3	48.3	Control 3 Parietal Ctx	16.8	45.7
Control 3 Temporal Ctx	30.6	51.4	Control (Path) 1 Parietal Ctx	82.4	37.4
Control 4 Temporal Ctx	17.4	33.2	Control (Path) 2 Parietal Ctx	49.3	58.6
Control (Path) 1 Temporal Ctx	70.7	21.2	Control (Path) 3 Parietal Ctx	14.4	0.3
Control (Path) 2 Temporal Ctx	44.8	32.1	Control (Path) 4 Parietal Ctx	71.7	7.6

Table YD. Panel 4.1D.

Tissue Name	Rel. Exp.(%)	Tissue Name	Rel. Exp.(%)
-------------	--------------	-------------	--------------

	Ag5696, Run 246509228		Ag5696, Run 246509228
Secondary Th1 act	0.9	HUVEC IL-1beta	7.6
Secondary Th2 act	0.2	HUVEC IFN gamma	12.8
Secondary Tr1 act	0.5	HUVEC TNF alpha + IFN gamma	0.9
Secondary Th1 rest	0.0	HUVEC TNF alpha + IL4	0.8
Secondary Th2 rest	0.0	HUVEC IL-11	8.1
Secondary Tr1 rest	0.0	Lung Microvascular EC none	17.1
Primary Th1 act	0.0	Lung Microvascular EC TNFalpha + IL-1beta	6.8
Primary Th2 act	0.3	Microvascular Dermal EC none	1.0
Primary Tr1 act	0.0	Microvascular Dermal EC TNFalpha + IL-1beta	3.4
Primary Th1 rest	0.0	Bronchial epithelium TNFalpha + IL1beta	4.5
Primary Th2 rest	0.3	Small airway epithelium none	5.7
Primary Tr1 rest	0.0	Small airway epithelium TNFalpha + IL-1beta	10.4
CD45RA CD4 lymphocyte act	25.9	Coronary artery SMC rest	21.5
CD45RO CD4 lymphocyte act	5.6	Coronary artery SMC TNFalpha + IL-1beta	20.7
CD8 lymphocyte act	0.6	Astrocytes rest	3.8
Secondary CD8 lymphocyte rest	3.7	Astrocytes TNFalpha + IL-1beta	2.0
Secondary CD8 lymphocyte act	0.3	KU-812 (Basophil) rest	7.8
CD4 lymphocyte none	0.4	KU-812 (Basophil) PMA/ionomycin	8.3
2ry. Th1/Th2/Tr1_anti- CD95 CH11	0.6	CCD1106 (Keratinocytes) none	37.4
LAK cells rest	3.0	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	11.0
LAK cells IL-2	2.6	Liver cirrhosis	14.8
LAK cells IL-2+IL-12	0.8	NCI-H292 none	44.1
LAK cells IL-2+IFN gamma	2.0	NCI-H292 IL-4	37.6
LAK cells IL-2+ IL-18	1.1	NCI-H292 IL-9	100.0
LAK cells	1.8	NCI-H292 IL-13	44.8

PMA/ionomycin			
NK Cells IL-2 rest	11.3	NCI-H292 IFN gamma	17.2
Two Way MLR 3 day	0.7	HPAEC none	4.2
Two Way MLR 5 day	0.0	HPAEC TNF alpha + IL-1 beta	33.0
Two Way MLR 7 day	0.4	Lung fibroblast none	79.6
PBMC rest	0.5	Lung fibroblast TNF alpha + IL-1 beta	48.3
PBMC PWM	0.2	Lung fibroblast IL-4	12.7
PBMC PHA-L	1.4	Lung fibroblast IL-9	37.1
Ramos (B cell) none	0.0	Lung fibroblast IL-13	6.3
Ramos (B cell) ionomycin	0.4	Lung fibroblast IFN gamma	37.6
B lymphocytes PWM	0.8	Dermal fibroblast CCD1070 rest	58.2
B lymphocytes CD40L and IL-4	0.3	Dermal fibroblast CCD1070 TNF alpha	46.0
EOL-1 dbcAMP	3.7	Dermal fibroblast CCD1070 IL-1 beta	39.2
EOL-1 dbcAMP PMA/ionomycin	0.3	Dermal fibroblast IFN gamma	28.1
Dendritic cells none	1.3	Dermal fibroblast IL-4	88.3
Dendritic cells LPS	0.0	Dermal Fibroblasts rest	35.1
Dendritic cells anti-CD40	0.0	Neutrophils TNFa+LPS	0.0
Monocytes rest	0.5	Neutrophils rest	0.3
Monocytes LPS	1.5	Colon	0.6
Macrophages rest	0.2	Lung	1.2
Macrophages LPS	0.4	Thymus	2.0
HUVEC none	5.7	Kidney	59.0
HUVEC starved	4.5		

**AI\_comprehensive panel\_v1.0 Summary:** Ag5696 Highest expression of this gene is seen in a normal bone sample adjacent to OA bone (CT=28). Overall, this gene is widely expressed on this panel, with moderate levels of expression in a wide range of tissues and samples related to autoimmune disease. Thus, modulation of the expression or function of this gene may be useful in the treatment of autoimmune diseases, including RA, OA, allergy, emphysema and asthma.

**CNS\_neurodegeneration\_v1.0 Summary:** Ag5696 Two experiments with the same probe and primer set produce results that are in very good agreement. This panel does not

- show differential expression of this gene in Alzheimer's disease. However, this panel does show that this gene is expressed at high to moderate levels in the hippocampus and cerebral cortex. Thus, therapeutic modulation of the expression or function of this gene may be useful in the treatment of neurological disorders, such as Alzheimer's disease, Parkinson's disease, schizophrenia, multiple sclerosis, stroke and epilepsy.

- Panel 4.1D Summary:** Ag5696 Highest expression of this gene is seen in IL-9 treated NCI-H292 goblet cells. Moderate levels of expression are seen in clusters of samples derived from lung and dermal fibroblasts. Low but significant levels of expression are seen in endothelial cells from the lung and skin, as well as small airway and bronchial epithelium. The prominent expression in cells and cell lines derived from the lung and skin suggest that this gene product may be involved in inflammatory conditions of the lung and skin, including psoriasis, asthma, emphysema, allergy, and chronic obstructive pulmonary disease.

**Z. CG155595-01: kinesin 7.**

- Expression of gene CG155595-01 was assessed using the primer-probe set Ag5284, described in Table ZA. Results of the RTQ-PCR runs are shown in Tables ZB, ZC, ZD and ZE.

Table ZA. Probe Name Ag5284

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-gatcagaggacctcgaggaa-3'	20	3979	546
Probe	TET-5'-ccacatgcacaaggattattccatacca-3'-TAMRA	28	3999	547
Reverse	5'-agaagctgcctgtctccttaat-3'	22	4043	548

Table ZB. AI\_comprehensive panel\_v1.0.

Tissue Name	Rel. Exp.(%) Ag5284, Run 234222219	Tissue Name	Rel. Exp.(%) Ag5284, Run 234222219
110967 COPD-F	8.5	112427 Match Control Psoriasis-F	100.0
110980 COPD-F	15.5	112418 Psoriasis-M	43.2
110968 COPD-M	17.8	112723 Match Control Psoriasis-M	14.6

110977 COPD-M	77.4	112419 Psoriasis-M	36.3
110989 Emphysema-F	38.4	112424 Match Control Psoriasis-M	23.2
110992 Emphysema-F	3.3	112420 Psoriasis-M	37.6
110993 Emphysema-F	16.8	112425 Match Control Psoriasis-M	66.9
110994 Emphysema-F	8.8	104689 (MF) OA Bone-Backus	23.8
110995 Emphysema-F	26.8	104690 (MF) Adj "Normal" Bone-Backus	19.2
110996 Emphysema-F	5.3	104691 (MF) OA Synovium-Backus	21.5
110997 Asthma-M	10.0	104692 (BA) OA Cartilage-Backus	14.4
111001 Asthma-F	5.7	104694 (BA) OA Bone-Backus	20.6
111002 Asthma-F	18.9	104695 (BA) Adj "Normal" Bone-Backus	10.3
111003 Atopic Asthma-F	18.8	104696 (BA) OA Synovium-Backus	9.5
111004 Atopic Asthma-F	22.1	104700 (SS) OA Bone-Backus	11.4
111005 Atopic Asthma-F	13.7	104701 (SS) Adj "Normal" Bone-Backus	6.0
111006 Atopic Asthma-F	2.8	104702 (SS) OA Synovium-Backus	14.8
111417 Allergy-M	2.0	117093 OA Cartilage Rep7	9.6
112347 Allergy-M	6.3	112672 OA Bone5	49.0
112349 Normal Lung-F	10.4	112673 OA Synovium5	20.3
112357 Normal Lung-F	87.7	112674 OA Synovial Fluid cells5	13.6
112354 Normal Lung-M	49.7	117100 OA Cartilage Rep14	2.0
112374 Crohns-F	21.0	112756 OA Bone9	29.7
112389 Match Control Crohns-F	15.6	112757 OA Synovium9	5.4
112375 Crohns-F	10.1	112758 OA Synovial Fluid Cells9	17.0
112732 Match	3.0	117125 RA Cartilage	8.7

Control Crohns-F		Rep2	
112725 Crohns-M	9.6	113492 Bone2 RA	4.7
112387 Match Control Crohns-M	3.1	113493 Synovium2 RA	0.0
112378 Crohns-M	15.2	113494 Syn Fluid Cells RA	5.9
112390 Match Control Crohns-M	73.2	113499 Cartilage4 RA	4.0
112726 Crohns-M	12.8	113500 Bone4 RA	16.8
112731 Match Control Crohns-M	32.1	113501 Synovium4 RA	2.5
112380 Ulcer Col-F	23.3	113502 Syn Fluid Cells4 RA	7.1
112734 Match Control Ulcer Col-F	21.3	113495 Cartilage3 RA	4.0
112384 Ulcer Col-F	33.9	113496 Bone3 RA	8.4
112737 Match Control Ulcer Col-F	9.0	113497 Synovium3 RA	0.0
112386 Ulcer Col-F	2.3	113498 Syn Fluid Cells3 RA	5.2
112738 Match Control Ulcer Col-F	6.5	117106 Normal Cartilage Rep20	5.1
112381 Ulcer Col-M	6.1	113663 Bone3 Normal	9.2
112735 Match Control Ulcer Col-M	34.2	113664 Synovium3 Normal	3.8
112382 Ulcer Col-M	23.8	113665 Syn Fluid Cells3 Normal	14.7
112394 Match Control Ulcer Col-M	3.4	117107 Normal Cartilage Rep22	0.0
112383 Ulcer Col-M	14.0	113667 Bone4 Normal	17.9
112736 Match Control Ulcer Col-M	8.9	113668 Synovium4 Normal	25.2
112423 Psoriasis-F	45.4	113669 Syn Fluid Cells4 Normal	24.7

Table ZC. CNS\_neurodegeneration\_v1.0

Tissue Name	Rel. Exp.(%) Ag5284, Run 233610763	Tissue Name	Rel. Exp.(%) Ag5284, Run 233610763
AD 1 Hippo	17.6	Control (Path) 3 Temporal Ctx	0.0
AD 2 Hippo	0.0	Control (Path) 4 Temporal Ctx	29.5
AD 3 Hippo	6.7	AD 1 Occipital	0.0

		Ctx	
AD 4 Hippo	0.0	AD 2 Occipital Ctx (Missing)	0.0
AD 5 hippo	47.0	AD 3 Occipital Ctx	6.9
AD 6 Hippo	19.6	AD 4 Occipital Ctx	8.8
Control 2 Hippo	7.0	AD 5 Occipital Ctx	6.3
Control 4 Hippo	15.7	AD 6 Occipital Ctx	12.2
Control (Path) 3 Hippo	6.7	Control 1 Occipital Ctx	5.9
AD 1 Temporal Ctx	0.0	Control 2 Occipital Ctx	35.1
AD 2 Temporal Ctx	26.6	Control 3 Occipital Ctx	42.0
AD 3 Temporal Ctx	4.8	Control 4 Occipital Ctx	0.0
AD 4 Temporal Ctx	19.1	Control (Path) 1 Occipital Ctx	10.3
AD 5 Inf Temporal Ctx	100.0	Control (Path) 2 Occipital Ctx	7.2
AD 5 SupTemporal Ctx	35.8	Control (Path) 3 Occipital Ctx	0.0
AD 6 Inf Temporal Ctx	15.7	Control (Path) 4 Occipital Ctx	15.6
AD 6 Sup Temporal Ctx	20.2	Control 1 Parietal Ctx	4.2
Control 1 Temporal Ctx	18.3	Control 2 Parietal Ctx	18.8
Control 2 Temporal Ctx	12.7	Control 3 Parietal Ctx	10.5
Control 3 Temporal Ctx	0.0	Control (Path) 1 Parietal Ctx	17.3
Control 4 Temporal Ctx	15.1	Control (Path) 2 Parietal Ctx	8.2
Control (Path) 1 Temporal Ctx	38.4	Control (Path) 3 Parietal Ctx	0.0
Control (Path) 2 Temporal Ctx	38.7	Control (Path) 4 Parietal Ctx	34.9

Table ZD. General\_screening\_panel\_v1.5

Tissue Name	Rel. Exp.(%)	Tissue Name	Rel. Exp.(%)
-------------	--------------	-------------	--------------

	Ag5284, Run 230564176		Ag5284, Run 230564176
Adipose	2.5	Renal ca. TK-10	23.7
Melanoma* Hs688(A).T	17.4	Bladder	6.1
Melanoma* Hs688(B).T	28.1	Gastric ca. (liver met.) NCI-N87	60.3
Melanoma* M14	32.8	Gastric ca. KATO III	36.9
Melanoma* LOXIMVI	23.3	Colon ca. SW-948	6.3
Melanoma* SK- MEL-5	18.0	Colon ca. SW480	41.2
Squamous cell carcinoma SCC-4	12.7	Colon ca.* (SW480 met) SW620	22.7
Testis Pool	1.6	Colon ca. HT29	10.4
Prostate ca.* (bone met) PC-3	9.5	Colon ca. HCT-116	100.0
Prostate Pool	1.5	Colon ca. CaCo-2	54.0
Placenta	0.5	Colon cancer tissue	8.3
Uterus Pool	2.2	Colon ca. SW1116	7.3
Ovarian ca. OVCAR-3	18.6	Colon ca. Colo-205	5.3
Ovarian ca. SK- OV-3	48.6	Colon ca. SW-48	5.7
Ovarian ca. OVCAR-4	11.3	Colon Pool	3.6
Ovarian ca. OVCAR-5	51.4	Small Intestine Pool	15.8
Ovarian ca. IGROV-1	8.4	Stomach Pool	3.7
Ovarian ca. OVCAR-8	15.8	Bone Marrow Pool	4.2
Ovary	4.2	Fetal Heart	5.4
Breast ca. MCF-7	19.3	Heart Pool	1.5
Breast ca. MDA- MB-231	37.9	Lymph Node Pool	12.2
Breast ca. BT 549	16.6	Fetal Skeletal Muscle	5.1
Breast ca. T47D	9.7	Skeletal Muscle Pool	0.4
Breast ca. MDA-N	24.7	Spleen Pool	2.6
Breast Pool	7.1	Thymus Pool	13.8
Trachea	1.4	CNS cancer (glio/astro) U87-MG	36.3
Lung	21.2	CNS cancer (glio/astro) U-118-MG	80.7

Fetal Lung	15.1	CNS cancer (neuro;met) SK-N-AS	46.3
Lung ca. NCI-N417	6.0	CNS cancer (astro) SF-539	12.0
Lung ca. LX-1	20.3	CNS cancer (astro) SNB-75	37.1
Lung ca. NCI-H146	2.8	CNS cancer (glio) SNB-19	5.1
Lung ca. SHP-77	44.1	CNS cancer (glio) SF-295	58.2
Lung ca. A549	46.7	Brain (Amygdala) Pool	0.3
Lung ca. NCI-H526	5.0	Brain (cerebellum)	0.3
Lung ca. NCI-H23	88.9	Brain (fetal)	10.4
Lung ca. NCI-H460	11.4	Brain (Hippocampus) Pool	0.6
Lung ca. HOP-62	13.4	Cerebral Cortex Pool	1.3
Lung ca. NCI-H522	30.4	Brain (Substantia nigra) Pool	0.6
Liver	0.0	Brain (Thalamus) Pool	2.3
Fetal Liver	24.0	Brain (whole)	1.5
Liver ca. HepG2	12.0	Spinal Cord Pool	1.9
Kidney Pool	24.1	Adrenal Gland	0.3
Fetal Kidney	45.7	Pituitary gland Pool	0.7
Renal ca. 786-0	18.3	Salivary Gland	0.5
Renal ca. A498	6.2	Thyroid (female)	1.4
Renal ca. ACHN	5.7	Pancreatic ca. CAPAN2	31.0
Renal ca. UO-31	7.5	Pancreas Pool	4.9

Table ZE. Panel 4.1D

Tissue Name	Rel. Exp.(%) Ag5284, Run 230510205	Tissue Name	Rel. Exp.(%) Ag5284, Run 230510205
Secondary Th1 act	37.9	HUVEC IL-1beta	14.6
Secondary Th2 act	40.6	HUVEC IFN gamma	18.8
Secondary Tr1 act	12.2	HUVEC TNF alpha + IFN gamma	6.6
Secondary Th1 rest	0.0	HUVEC TNF alpha + IL4	5.3
Secondary Th2 rest	2.1	HUVEC IL-11	3.2
Secondary Tr1 rest	7.7	Lung Microvascular EC none	17.0
Primary Th1 act	5.4	Lung Microvascular EC TNFalpha + IL-1beta	1.7

Primary Th2 act	12.7	Microvascular Dermal EC none	8.7
Primary Tr1 act	13.1	Microvasular Dermal EC TNFalpha + IL-1beta	1.3
Primary Th1 rest	0.0	Bronchial epithelium TNFalpha + IL1beta	0.0
Primary Th2 rest	6.5	Small airway epithelium. none	0.0
Primary Tr1 rest	6.0	Small airway epithelium TNFalpha + IL-1beta	8.7
CD45RA CD4 lymphocyte act	40.3	Coronery artery SMC rest	0.0
CD45RO CD4 lymphocyte act	31.9	Coronery artery SMC TNFalpha + IL-1beta	4.8
CD8 lymphocyte act	19.5	Astrocytes rest	4.1
Secondary CD8 lymphocyte rest	12.2	Astrocytes TNFalpha + IL-1beta	3.7
Secondary CD8 lymphocyte act	0.0	KU-812 (Basophil) rest	33.9
CD4 lymphocyte none	0.0	KU-812 (Basophil) PMA/ionomycin	37.4
2ry Th1/Th2/Tr1_anti- CD95 CH11	0.0	CCD1106 (Keratinocytes) none	31.9
LAK cells rest	1.2	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	5.0
LAK cells IL-2	13.0	Liver cirrhosis	3.9
LAK cells IL-2+IL-12	2.2	NCI-H292 none	36.6
LAK cells IL-2+IFN gamma	9.3	NCI-H292 IL-4	46.0
LAK cells IL-2+ IL-18	2.2	NCI-H292 IL-9	73.2
LAK cells PMA/ionomycin	1.9	NCI-H292 IL-13	72.7
NK Cells IL-2 rest	47.6	NCI-H292 IFN gamma	28.1
Two Way MLR 3 day	3.4	HPAEC none	2.8
Two Way MLR 5 day	2.5	HPAEC TNF alpha + IL- 1 beta	11.1
Two Way MLR 7 day	9.4	Lung fibroblast none	9.2
PBMC rest	0.0	Lung fibroblast TNF alpha + IL-1 beta	7.0
PBMC PWM	3.3	Lung fibroblast IL-4	3.4
PBMC PHA-L	19.8	Lung fibroblast IL-9	11.8
Ramos (B cell) none	11.9	Lung fibroblast IL-13	1.3
Ramos (B cell)	17.8	Lung fibroblast IFN	5.5

ionomycin		gamma	
B lymphocytes PWM	13.7	Dermal fibroblast CCD1070 rest	20.9
B lymphocytes CD40L and IL-4	18.3	Dermal fibroblast CCD1070 TNF alpha	100.0
EOL-1 dbcAMP	24.0	Dermal fibroblast CCD1070 IL-1 beta	24.1
EOL-1 dbcAMP PMA/ionomycin	21.6	Dermal fibroblast IFN gamma	12.3
Dendritic cells none	1.6	Dermal fibroblast IL-4	38.7
Dendritic cells LPS	0.0	Dermal Fibroblasts rest	7.2
Dendritic cells anti- CD40	0.0	Neutrophils TNFa+LPS	0.0
Monocytes rest	0.0	Neutrophils rest	0.0
Monocytes LPS	0.0	Colon	0.0
Macrophages rest	0.0	Lung	0.0
Macrophages LPS	0.0	Thymus	4.0
HUVEC none	3.1	Kidney	0.0
HUVEC starved	22.7		

**AI\_comprehensive\_panel\_v1.0 Summary:** Ag5284 Highest expression of this gene is seen in a normal tissue sample adjacent to psoriatic tissue (CT=33).

**CNS\_neurodegeneration\_v1.0 Summary:** Ag5284 Expression is limited to a single inferior temporal cortex sample from an Alzheimer's patient (CT=34.9).

5. **General\_screening\_panel\_v1.5 Summary:** Ag5284 Highest expression is seen in a colon cancer cell line (CT=31). Prominent levels of expression are also seen in cell lines derived from brain, lung, colon, gastric, pancreatic, breast, ovarian, and melanoma cancers. Thus, expression of this gene could be used as a marker to detect the presence of these cancers. Furthermore, therapeutic modulation of the expression or function of this gene may be
10. effective in the treatment of brain, lung, colon, gastric, pancreatic, breast, ovarian, and melanoma cancers.

15. **Panel 4.1D Summary:** Ag5284 Highest expression of this gene is seen in TNF-a treated dermal fibroblasts (CT=33). Low but significant levels of expression are also seen in clusters of samples derived from basophils, NCI-H292 cells, resting NK cells, and secondary activated T cells.

**AA. CG157477-01: MYOSIN I.**

Expression of gene CG157477-01 was assessed using the primer-probe set Ag5289, described in Table AAA. Results of the RTQ-PCR runs are shown in Tables AAB, AAC and AAD.

5 Table AAA. Probe Name Ag5289

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5' - cgcacatctatacgttcattgga - 3'	21	151	549
Probe	TET-5' - tcgtcggtttctgtgaacccttacaag - 3' - TAMRA	26	176	550
Reverse	5' - tgctcaattgtgtctcttccat - 3'	22	215	551

Table AAB. CNS\_neurodegeneration\_v1.0

Tissue Name	Rel. Exp.(%) Ag5289, Run 233610765	Tissue Name	Rel. Exp.(%) Ag5289, Run 233610765
AD 1 Hippo	14.0	Control (Path) 3 Temporal Ctx	3.9
AD 2 Hippo	29.9	Control (Path) 4 Temporal Ctx	28.1
AD 3 Hippo	12.9	AD 1 Occipital Ctx	24.8
AD 4 Hippo	12.5	AD 2 Occipital Ctx (Missing)	0.0
AD 5 Hippo	49.0	AD 3 Occipital Ctx	11.5
AD 6 Hippo	42.9	AD 4 Occipital Ctx	25.2
Control 2 Hippo	37.1	AD 5 Occipital Ctx	44.1
Control 4 Hippo	24.1	AD 6 Occipital Ctx	22.5
Control (Path) 3 Hippo	10.7	Control 1 Occipital Ctx	8.1
AD 1 Temporal Ctx	36.3	Control 2 Occipital Ctx	49.7
AD 2 Temporal Ctx	37.9	Control 3 Occipital Ctx	19.9
AD 3 Temporal Ctx	10.4	Control 4 Occipital Ctx	15.8
AD 4 Temporal Ctx	29.7	Control (Path) 1 Occipital Ctx	100.0
AD 5 Inf Temporal Ctx	83.5	Control (Path) 2 Occipital Ctx	25.5
AD 5 Sup Temporal Ctx	36.1	Control (Path) 3 Occipital Ctx	4.2

AD 6 Inf Temporal Ctx	61.1	Control (Path) 4 Occipital Ctx	20.3
AD 6 Sup Temporal Ctx	47.0	Control 1 Parietal Ctx	17.3
Control 1 Temporal Ctx	7.7	Control 2 Parietal Ctx	39.0
Control 2 Temporal Ctx	38.7	Control 3 Parietal Ctx	21.5
Control 3 Temporal Ctx	18.8	Control (Path) 1 Parietal Ctx	50.0
Control 3 Temporal Ctx	9.2	Control (Path) 2 Parietal Ctx	39.5
Control (Path) 1 Temporal Ctx	53.6	Control (Path) 3 Parietal Ctx	4.1
Control (Path) 2 Temporal Ctx	32.5	Control (Path) 4 Parietal Ctx	38.2

Table AAC. General\_screening\_panel\_v1.5

Tissue Name	Rel. Exp.(%) Ag5289, Run 233238980	Tissue Name	Rel. Exp.(%) Ag5289, Run 233238980
Adipose	7.2	Renal ca. TK-10	14.0
Melanoma* Hs688(A).T	65.1	Bladder	19.6
Melanoma* Hs688(B).T	16.2	Gastric ca. (liver met.) NCI-N87	21.3
Melanoma* M14	23.3	Gastric ca. KATO III	50.3
Melanoma* LOXIMVI	8.1	Colon ca. SW-948	1.5
Melanoma* SK- MEL-5	11.2	Colon ca. SW480	100.0
Squamous cell carcinoma SCC-4	3.1	Colon ca.* (SW480 met) SW620	12.9
Testis Pool	4.0	Colon ca. HT29	9.5
Prostate ca.* (bone met) PC-3	28.7	Colon ca. HCT-116	11.8
Prostate Pool	7.4	Colon ca. CaCo-2	66.9
Placenta	5.9	Colon cancer tissue	19.5
Uterus Pool	9.7	Colon ca. SW1116	3.4
Ovarian ca. OVCAR-3	2.1	Colon ca. Colo-205	3.2
Ovarian ca. SK- OV-3	17.3	Colon ca. SW-48	11.6

Ovarian ca. OVCAR-4	6.0	Colon Pool	9.0
Ovarian ca. OVCAR-5	34.9	Small Intestine Pool	6.3
Ovarian ca. IGROV-1	1.5	Stomach Pool	3.7
Ovarian ca. OVCAR-8	1.6	Bone Marrow Pool	5.3
Ovary	5.6	Fetal Heart	1.2
Breast ca. MCF-7	11.6	Heart Pool	3.6
Breast ca. MDA- MB-231	0.5	Lymph Node Pool	10.4
Breast ca. BT 549	0.1	Fetal Skeletal Muscle	0.7
Breast ca. T47D	17.6	Skeletal Muscle Pool	2.4
Breast ca. MDA-N	4.4	Spleen Pool	5.7
Breast Pool	8.5	Thymus Pool	5.8
Trachea	17.6	CNS cancer (glio/astro) U87-MG	5.6
Lung	3.1	CNS cancer (glio/astro) U-118-MG	1.5
Fetal Lung	15.4	CNS cancer (neuro;met) SK-N-AS	0.2
Lung ca. NCI-N417	1.8	CNS cancer (astro) SF- 539	0.2
Lung ca. LX-1	34.2	CNS cancer (astro) SNB-75	0.1
Lung ca. NCI-H146	8.2	CNS cancer (glio) SNB-19	1.2
Lung ca. SHP-77	5.6	CNS cancer (glio) SF- 295	0.6
Lung ca. A549	2.6	Brain (Amygdala) Pool	6.3
Lung ca. NCI-H526	2.0	Brain (cerebellum)	11.0
Lung ca. NCI-H23	1.7	Brain (fetal)	4.5
Lung ca. NCI-H460	0.7	Brain (Hippocampus) Pool	6.2
Lung ca. HOP-62	1.6	Cerebral Cortex Pool	7.3
Lung ca. NCI-H522	0.6	Brain (Substantia nigra) Pool	4.7
Liver	0.9	Brain (Thalamus) Pool	7.7
Fetal Liver	10.4	Brain (whole)	6.4
Liver ca. HepG2	13.3	Spinal Cord Pool	12.2
Kidney Pool	15.0	Adrenal Gland	15.0
Fetal Kidney	4.9	Pituitary gland Pool	1.8

Renal ca. 786-0	1.5	Salivary Gland	5.4
Renal ca. A498	2.2	Thyroid (female)	7.0
Renal ca. ACHN	28.1	Pancreatic ca. CAPAN2	27.0
Renal ca. UO-31	7.0	Pancreas Pool	8.7

Table AAD. Panel 4.1D

Tissue Name	Rel. Exp.(%) Ag5289, Run 233229299	Rel. Exp.(%) Ag5289, Run 233232664	Tissue Name	Rel. Exp.(%) Ag5289, Run 233229299	Rel. Exp.(%) Ag5289, Run 233232664
Secondary Th1 act	0.9	0.9	HUVEC IL-1beta	12.2	24.0
Secondary Th2 act	1.3	1.9	HUVEC IFN gamma	12.8	16.6
Secondary Tr1 act	0.1	0.7	HUVEC TNF alpha + IFN gamma	1.3	2.0
Secondary Th1 rest	0.0	0.0	HUVEC TNF alpha + IL4	3.2	3.8
Secondary Th2 rest	0.0	0.0	HUVEC IL-11	7.4	12.7
Secondary Tr1 rest	0.0	0.0	Lung Microvascular EC none	41.2	65.5
Primary Th1 act	0.0	0.0	Lung Microvascular EC TNFalpha + IL- 1beta	9.9	13.5
Primary Th2 act	0.5	0.9	Microvascular Dermal EC none	0.8	1.2
Primary Tr1 act	0.3	0.6	Microvascular Dermal EC TNFalpha + IL- 1beta	3.1	4.1
Primary Th1 rest	0.0	0.0	Bronchial epithelium TNFalpha + IL1beta	10.6	27.4
Primary Th2 rest	0.0	0.1	Small airway epithelium none	7.3	13.1
Primary Tr1 rest	0.0	0.0	Small airway epithelium TNFalpha + IL- 1beta	15.5	27.4
CD45RA CD4	5.1	4.9	Coronary artery	1.3	2.1

lymphocyte act			SMC rest		
CD45RO CD4 lymphocyte act	2.4	4.2	Coronary artery SMC TNFalpha + IL-1beta	1.8	2.0
CD8 lymphocyte act	0.3	0.5	Astrocytes rest	0.1	0.1
Secondary CD8 lymphocyte rest	1.8	2.5	Astrocytes TNFalpha + IL-1beta	0.1	0.2
Secondary CD8 lymphocyte act	0.0	0.1	KU-812 (Basophil) rest	6.4	11.8
CD4 lymphocyte none	0.0	0.0	KU-812 (Basophil) PMA/ionomycin	20.2	35.8
2ry Th1/Th2/Tr1_anti-CD95 CH11	0.0	0.0	CCD1106 (Keratinocytes) none	100.0	13.4
LAK cells rest	0.7	0.9	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	8.2	14.0
LAK cells IL-2	0.6	0.9	Liver cirrhosis	3.4	5.3
LAK cells IL-2+IL-12	0.1	0.2	NCI-H292 none	6.7	15.3
LAK cells IL-2+IFN gamma	0.5	0.9	NCI-H292 IL-4	8.8	13.1
LAK cells IL-2+IL-18	0.3	0.3	NCI-H292 IL-9	13.7	32.1
LAK cells PMA/ionomycin	2.5	4.3	NCI-H292 IL-13	12.4	15.6
NK Cells IL-2 rest	4.3	4.3	NCI-H292 IFN gamma	3.9	7.6
Two Way MLR 3 day	0.5	0.5	HPAEC none	3.4	4.6
Two Way MLR 5 day	0.1	0.0	HPAEC TNF alpha + IL-1 beta	11.3	16.2
Two Way MLR 7 day	0.2	0.4	Lung fibroblast none	1.2	1.7
PBMC rest	0.1	0.2	Lung fibroblast TNF alpha + IL-1 beta	0.1	0.5
PBMC PWM	0.2	0.4	Lung fibroblast IL-4	1.9	4.2
PBMC PHA-L	1.0	1.0	Lung fibroblast IL-9	1.7	2.1

Ramos (B cell) none	1.3	2.6	Lung fibroblast IL-13	0.2	0.5
Ramos (B cell) ionomycin	26.1	29.3	Lung fibroblast IFN gamma	1.9	1.9
B lymphocytes PWM	1.3	2.4	Dermal fibroblast CCD1070.rest	3.9	6.7
B lymphocytes CD40L and IL-4	5.4	8.8	Dermal fibroblast CCD1070.TNF alpha	4.4	7.5
EOL-1 dbcAMP	0.0	0.0	Dermal fibroblast CCD1070.IL-1 beta	3.9	7.4
EOL-1 dbcAMP. PMA/ionomycin	0.0	0.0	Dermal fibroblast IFN gamma	12.2	21.9
Dendritic cells none	0.5	1.1	Dermal fibroblast IL-4	72.2	100.0
Dendritic cells LPS	0.0	0.0	Dermal Fibroblasts rest	10.7	18.8
Dendritic cells anti-CD40	0.1	0.3	Neutrophils TNFa+LPS	0.1	0.2
Monocytes rest	0.0	0.0	Neutrophils rest	0.1	0.0
Monocytes LPS	0.3	0.7	Colon	1.6	4.0
Macrophages rest	0.5	0.4	Lung	1.0	2.2
Macrophages LPS	0.5	0.9	Thymus	0.6	0.5
HUVEC none	8.5	10.5	Kidney	4.7	6.3
HUVEC starved	17.7	26.2			

**CNS\_neurodegeneration\_v1.0 Summary:** Ag5289 This panel does not show differential expression of this gene in Alzheimer's disease. However, this profile confirms the expression of this gene at moderate levels in the brain. Please see Panel 1.5 for discussion of utility of this gene in the central nervous system.

5. **General\_screening\_panel\_v1.5 Summary:** Ag5289 Highest expression of this gene is seen in a colon cancer cell line (CT=23.5). This gene is widely expressed in this panel, with high levels of expression seen in brain, colon, gastric, lung, breast, ovarian, and melanoma cancer cell lines. This expression profile suggests a role for this gene product in cell survival and proliferation. Modulation of this gene product may be useful in the treatment
- 10 of cancer.

Among tissues with metabolic function, this gene is expressed at high to moderate levels in pituitary, adipose, adrenal gland, pancreas, thyroid, and adult and fetal skeletal muscle,

heart, and liver. This widespread expression among these tissues suggests that this gene product may play a role in normal neuroendocrine and metabolic function and that dysregulated expression of this gene may contribute to neuroendocrine disorders or metabolic diseases, such as obesity and diabetes.

- 5 In addition, this gene is expressed at much higher levels in fetal liver tissue (CT=26.7) when compared to expression in the adult counterpart (CT=30.3). Thus, expression of this gene may be used to differentiate between the fetal and adult source of this tissue.

This gene is also expressed at high levels in the CNS, including the hippocampus, thalamus, substantia nigra, amygdala, cerebellum and cerebral cortex. Therefore,  
10 therapeutic modulation of the expression or function of this gene may be useful in the treatment of neurologic disorders, such as Alzheimer's disease, Parkinson's disease, schizophrenia, multiple sclerosis, stroke and epilepsy.

**Panel 4.1D Summary:** Ag5289 Highest expression is seen in IL-4 treated dermal fibroblasts (CT=26.5). Moderate levels of expression are also seen in clusters of samples  
15 derived from lung and dermal fibroblasts, endothelial cells from lung, skin, umbilical vein, and pulmonary artery, small airway and bronchial epithelial cells, and NCI-H292 mucoc-epidermoid cells. The preponderance of expression in cells derived from the lung and skin suggests that this gene product may be involved in inflammatory processes that involve these organs. Therefore, therapeutic modulation of the expression or function of this gene  
20 product may be useful in the treatment of psoriasis, asthma, allergy, and emphysema. A second run with the same probe and primer set, run 233229299, is not included because the amp plot indicates there were experimental difficulties with this run.

**AB. CG157486-01: Ephrin receptor A2.**

Expression of gene CG157486-01 was assessed using the primer-probe set Ag2620,  
25 described in Table ABA. Results of the RTQ-PCR runs are shown in Tables ABB, ABC, ABD, ABE and ABF.

Table ABA. Probe Name Ag2620

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-gaagtgggtactgctggactttg-3'.	22	195	552

Probe	TET-5'-ctcacacacccgtatggcaaagggt-3'- TAMRA	25	243	553
Reverse	5'-cattcatgatgttctgcatcag-3'	22	273	554

Table ABB. General\_screening\_panel\_v1.5

Tissue Name	Rel. Exp.(%) Ag2620, Run 229827540	Tissue Name	Rel. Exp.(%) Ag2620, Run 229827540
Adipose	2.4	Renal ca. TK-10	29.1
Melanoma* Hs688(A).T	6.1	Bladder	3.7
Melanoma* Hs688(B).T	7.7	Gastric ca. (liver met.) NCI-N87	69.7
Melanoma* M14	0.7	Gastric ca. KATO III	69.3
Melanoma* LOXIMVI	12.7	Colon ca. SW-948	23.8
Melanoma* SK- MEL-5	1.8	Colon ca. SW480	36.9
Squamous cell carcinoma SCC-4	11.0	Colon ca.* (SW480 met) SW620	22.5
Testis Pool	0.4	Colon ca. HT29	7.9
Prostate ca.* (bone met) PC-3	100.0	Colon ca. HCT-116	30.8
Prostate Pool	0.7	Colon ca. CaCo-2	6.1
Placenta	2.4	Colon cancer tissue	13.8
Uterus Pool	1.8	Colon ca. SW1116	4.2
Ovarian ca. OVCAR-3	25.5	Colon ca. Colo-205	1.7
Ovarian ca. SK- OV-3	64.6	Colon ca. SW-48	5.3
Ovarian ca. OVCAR-4	17.0	Colon Pool	2.6
Ovarian ca. OVCAR-5	37.4	Small Intestine Pool	1.4
Ovarian ca. IGROV-1	41.8	Stomach Pool	1.9
Ovarian ca. OVCAR-8	18.6	Bone Marrow Pool	0.4
Ovary	1.2	Fetal Heart	0.7
Breast ca. MCF-7	2.5	Heart Pool	1.1
Breast ca. MDA- MB-231	57.4	Lymph Node Pool	1.2
Breast ca. BT 549	22.8	Fetal Skeletal Muscle	0.3

Breast ca. T47D	0.2	Skeletal Muscle Pool	1.1
Breast ca. MDA-N	0.9	Spleen Pool	2.1
Breast Pool	1.5	Thymus Pool	0.9
Trachea	4.2	CNS cancer (glio/astro) U87-MG	1.0
Lung	0.0	CNS cancer (glio/astro) U-118-MG	19.5
Fetal Lung	7.3	CNS cancer (neuro;met) SK-N-AS	7.2
Lung ca. NCI-N417	0.7	CNS cancer (astro) SF-539	12.3
Lung ca. LX-1	40.3	CNS cancer (astro) SNB-75	23.2
Lung ca. NCI-H146	0.1	CNS cancer (glio) SNB-19	41.8
Lung ca. SHP-77	0.3	CNS cancer (glio) SF-295	42.9
Lung ca. A549	9.6	Brain (Amygdala) Pool	0.1
Lung ca. NCI-H526	0.4	Brain (cerebellum)	0.3
Lung ca. NCI-H23	4.8	Brain (fetal)	0.5
Lung ca. NCI-H460	5.2	Brain (Hippocampus) Pool	0.2
Lung ca. HOP-62	22.4	Cerebral Cortex Pool	0.1
Lung ca. NCI-H522	12.0	Brain (Substantia nigra) Pool	0.4
Liver	0.4	Brain (Thalamus) Pool	0.3
Fetal Liver	1.1	Brain (whole)	0.2
Liver ca. HepG2	19.2	Spinal Cord Pool	0.4
Kidney Pool	5.1	Adrenal Gland	1.4
Fetal Kidney	1.2	Pituitary gland Pool	0.1
Renal ca. 786-0	18.0	Salivary Gland	7.1
Renal ca. A498	3.5	Thyroid (female)	2.7
Renal ca. ACHN	12.3	Pancreatic ca. CAPAN2	59.5
Renal ca. UO-31	22.5	Pancreas Pool	2.1

Table ABC. Oncology\_cell\_line\_screening\_panel\_v3.1

Tissue Name	Rel. Exp.(%) Ag2620, Run 230277126	Tissue Name	Rel. Exp.(%) Ag2620, Run 230277126
Daoy	1.5	Ca Ski Cervical epidermoid	58.2

Medulloblastoma/Cerebellum		carcinoma (metastasis)	
TE671 Medulloblastom/Cerebellum	3.1	ES-2_Ovarian clear cell carcinoma	15.8
D283 Med Medulloblastoma/Cerebellum	24.5	Ramos/6h stim_ Stimulated with PMA/ionomycin 6h	0.0
PFSK-1 Primitive Neuroectodermal/Cerebellum	19.3	Ramos/14h stim_ Stimulated with PMA/ionomycin 14h	0.0
XF-498_CNS	23.5	MEG-01_Chronic myelogenous leukemia (megokaryoblast)	0.2
SNB-78_CNS/glioma	5.5	Raji_Burkitt's lymphoma	0.1
SF-268_CNS/glioblastoma	29.3	Daudi_Burkitt's lymphoma	0.0
T98G_Glioblastoma	13.6	U266_B-cell plasmacytoma/myeloma	0.0
SK-N-SH_Neuroblastoma (metastasis)	6.5	CA46_Burkitt's lymphoma	0.0
SF-295_CNS/glioblastoma	17.3	RL_non-Hodgkin's B-cell lymphoma	0.0
Cerebellum	0.1	JM1_pre-B-cell lymphoma/leukemia	0.0
Cerebellum	0.0	Jurkat_T cell leukemia	0.0
NCI-H292_Mucoepidermoid lung ca.	83.5	TF-1_Erythroleukemia	0.1
DMS-114_Small cell lung cancer	3.3	HUT 78_T-cell lymphoma	0.7
DMS-79_Small cell lung cancer/neuroendocrine	0.9	U937_Histiocytic lymphoma	0.0
NCI-H146_Small cell lung cancer/neuroendocrine	0.4	KU-812_Myelogenous leukemia	0.0
NCI-H526_Small cell lung cancer/neuroendocrine	1.0	769-P_Clear cell renal ca.	9.3
NCI-N417_Small cell lung cancer/neuroendocrine	0.6	Caki-2_Clear cell renal ca.	9.9
NCI-H82_Small cell lung cancer/neuroendocrine	0.7	SW 839_Clear cell renal ca.	31.2
NCI-H157_Squamous cell lung cancer (metastasis)	14.0	G401_Wilms' tumor	4.6
NCI-H1155_Large cell lung cancer/neuroendocrine	0.1	Hs766T_Pancreatic ca. (LN metastasis)	100.0
NCI-H1299_Large cell lung cancer/neuroendocrine	21.9	CAPAN-1_Pancreatic adenocarcinoma (liver metastasis)	50.0
NCI-H727_Lung carcinoid	14.5	SU86.86_Pancreatic carcinoma (liver metastasis)	64.2
NCI-UMC-11_Lung carcinoid	0.0	BxPC-3_Pancreatic adenocarcinoma	35.1

LX-1_Small cell lung cancer	20.3	HPAC_Pancreatic adenocarcinoma	58.6
Colo-205_Colon cancer	1.9	MIA PaCa-2_Pancreatic ca.	18.3
KM12_Colon cancer	16.3	CFPAC-1_Pancreatic ductal adenocarcinoma	73.7
KM20L2_Colon cancer	9.5	PANC-1_Pancreatic epithelioid ductal ca.	70.2
NCI-H716_Colon cancer	15.1	T24_Bladder ca. (transitional cell)	16.5
SW-48_Colon adenocarcinoma	5.2	5637_Bladder ca.	35.8
SW1116_Colon adenocarcinoma	5.0	HT-1197_Bladder ca.	35.1
LS 174T_Colon adenocarcinoma	25.2	UM-UC-3_Bladder ca. (transitional cell)	9.3
SW-948_Colon adenocarcinoma	1.4	A204_Rhabdomyosarcoma	6.7
SW-480_Colon adenocarcinoma	3.3	HT-1080_Fibrosarcoma	18.0
NCI-SNU-5_Gastric ca.	14.7	MG-63_Osteosarcoma (bone)	11.3
KATO III_Stomach	20.7	SK-LMS-1_Leiomyosarcoma (vulva)	12.9
NCI-SNU-16_Gastric ca.	8.8	SJRH30_Rhabdomyosarcoma (met to bone marrow)	12.2
NCI-SNU-1_Gastric ca.	6.1	A431_Epidermoid ca.	36.6
RF-1_Gastric adenocarcinoma	0.1	WM266-4_Melanoma	0.3
RF-48_Gastric adenocarcinoma	0.1	DU 145_Prostate	12.3
MKN-45_Gastric ca.	27.5	MDA-MB-468_Breast adenocarcinoma	2.7
NCI-N87_Gastric ca.	20.0	SSC-4_Tongue	7.5
OVCAR-5_Ovarian ca.	16.2	SSC-9_Tongue	12.2
RL95-2_Uterine carcinoma	4.2	SSC-15_Tongue	9.3
HelaS3_Cervical adenocarcinoma	9.0	CAL 27_Squamous cell ca. of tongue	17.0

Table ABD. Panel 1.3D

Tissue Name	Rel. Exp.(%) Ag2620, Run 167660097	Tissue Name	Rel. Exp.(%) Ag2620, Run 167660097
Liver adenocarcinoma	52.9	Kidney (fetal)	26.6
Pancreas	2.6	Renal ca. 786-0	21.0
Pancreatic ca. CAPAN	33.0	Renal ca. A498	30.6

2			
Adrenal gland	0.9	Renal ca. RXF 393	29.3
Thyroid	0.6	Renal ca. ACHN	25.0
Salivary gland	8.8	Renal ca. UO-31	17.2
Pituitary gland	0.5	Renal ca. TK-10	20.7
Brain (fetal)	1.7	Liver	0.7
Brain (whole)	0.3	Liver (fetal)	3.5
Brain (amygdala)	0.7	Liver ca. (hepatoblast) HepG2	17.4
Brain (cerebellum)	0.0	Lung	3.3
Brain (hippocampus)	1.0	Lung (fetal)	3.0
Brain (substantia nigra)	0.9	Lung ca. (small cell) LX-1	21.6
Brain (thalamus)	0.6	Lung ca. (small cell) NCI-H69	0.0
Cerebral Cortex	0.4	Lung ca. (s.cell var.) SHP-77	0.9
Spinal cord	1.5	Lung ca. (large cell)NCI-H460	1.8
glio/astro U87-MG	1.3	Lung ca. (non-sm. cell) A549	8.8
glio/astro U-118-MG	14.1	Lung ca. (non-s.cell) NCI-H23	3.9
astrocytoma SW1783	25.5	Lung ca. (non-s.cell) HOP-62	28.3
neuro*; met SK-N-AS	3.7	Lung ca. (non-s.cl) NCI-H522	16.7
astrocytoma SF-539	9.0	Lung ca. (squam.) SW 900	15.5
astrocytoma SNB-75	21.3	Lung ca. (squam.) NCI-H596	0.2
glioma SNB-19	21.0	Mammary gland	5.1
glioma U251	35.1	Breast ca.* (pl.ef) MCF-7	1.5
glioma SF-295	31.6	Breast ca.* (pl.ef) MDA-MB-231	41.8
Heart (fetal)	16.6	Breast ca.* (pl.ef) T47D	0.5
Heart	1.2	Breast ca. BT-549	28.7
Skeletal muscle (fetal)	2.7	Breast ca. MDA-N	1.1
Skeletal muscle	0.7	Ovary	2.3
Bone marrow	0.3	Ovarian ca. OVCAR-3	33.0
Thymus	1.0	Ovarian ca.	18.9

		OVCAR-4	
Spleen	1.5	Ovarian ca. OVCAR-5	92.0
Lymph node	4.2	Ovarian ca. OVCAR-8	3.4
Colorectal	4.4	Ovarian ca. IGROV-1	5.0
Stomach	1.0	Ovarian ca.* (ascites) SK-OV-3	100.0
Small intestine	1.6	Uterus	2.1
Colon ca. SW480	27.2	Placenta	2.4
Colon ca.* SW620(SW480 met)	39.8	Prostate	1.2
Colon ca. HT29	9.5	Prostate ca.* (bone met)PC-3	64.6
Colon ca. HCT-116	14.0	Testis	0.4
Colon ca. CaCo-2	7.1	Melanoma Hs688(A).T	4.1
Colon ca. tissue(ODO3866)	13.3	Melanoma* (met) Hs688(B).T	3.9
Colon ca. HCC-2998	49.7	Melanoma UACC-62	6.3
Gastric ca.* (liver met) NCI-N87	48.3	Melanoma M14	0.0
Bladder	1.9	Melanoma LOX IMVI	14.0
Trachea	4.3	Melanoma* (met) SK-MEL-5	0.9
Kidney	3.3	Adipose	7.0

Table ABE. Panel 2.2

Tissue Name	Rel. Exp.(%) Ag2620, Run 175135887	Tissue Name	Rel. Exp.(%) Ag2620, Run 175135887
Normal Colon	6.9	Kidney Margin (OD04348)	100.0
Colon cancer (OD06064)	34.9	Kidney malignant cancer (OD06204B)	11.6
Colon Margin (OD06064)	3.7	Kidney normal adjacent tissue (OD06204E)	3.4
Colon cancer (OD06159)	18.9	Kidney Cancer (OD04450-01)	87.7
Colon Margin	1.9	Kidney Margin	5.1

(OD06159)		(OD04450-03)	
Colon cancer (OD06297-04)	9.3	Kidney Cancer 8120613	0.0
Colon Margin (OD06297-05)	14.5	Kidney Margin 8120614	5.4
CC Gr.2 ascend colon (ODO3921)	38.2	Kidney Cancer 9010320	17.6
CC Margin (ODO3921)	8.8	Kidney Margin 9010321	8.2
Colon cancer metastasis (OD06104)	1.7	Kidney Cancer 8120607	42.3
Lung Margin (OD06104)	3.0	Kidney Margin 8120608	18.7
Colon mets to lung (OD04451-01)	28.9	Normal Uterus	11.0
Lung Margin (OD04451-02)	6.3	Uterine Cancer 064011	11.5
Normal Prostate	3.0	Normal Thyroid	2.0
Prostate Cancer (OD04410)	1.4	Thyroid Cancer 064010	46.3
Prostate Margin (OD04410)	1.6	Thyroid Cancer A302152	20.2
Normal Ovary	12.1	Thyroid Margin A302153	9.9
Ovarian cancer (OD06283-03)	2.7	Normal Breast	12.9
Ovarian Margin (OD06283-07)	5.5	Breast Cancer (OD04566)	1.2
Ovarian Cancer 064008	16.3	Breast Cancer 1024	5.8
Ovarian cancer (OD06145)	10.4	Breast Cancer (OD04590-01)	0.2
Ovarian Margin (OD06145)	8.4	Breast Cancer Mets (OD04590-03)	2.4
Ovarian cancer (OD06455-03)	22.7	Breast Cancer Metastasis (OD04655-05)	16.3
Ovarian Margin (OD06455-07)	2.8	Breast Cancer 064006	1.6
Normal Lung	7.0	Breast Cancer 9100266	5.2
Invasive poor diff. lung adeno (ODO4945-01)	1.6	Breast Margin 9100265	2.5
Lung Margin (ODO4945-03)	25.3	Breast Cancer A209073	4.5
Lung Malignant Cancer (OD03126)	3.3	Breast Margin A2090734	14.3

Lung Margin (OD03126)	16.2	Breast cancer (OD06083)	3.9
Lung Cancer (OD05014A)	22.4	Breast cancer node metastasis (OD06083)	2.2
Lung Margin (OD05014B)	15.5	Normal Liver	7.9
Lung cancer (OD06081)	5.6	Liver Cancer 1026	19.3
Lung Margin (OD06081)	2.9	Liver Cancer 1025	18.2
Lung Cancer (OD04237-01)	13.3	Liver Cancer 6004-T	12.9
Lung Margin (OD04237-02)	37.1	Liver Tissue 6004-N	3.7
Ocular Melanoma Metastasis	11.3	Liver Cancer 6005-T	11.3
Ocular Melanoma Margin (Liver)	35.8	Liver Tissue 6005-N	28.1
Melanoma Metastasis	7.3	Liver Cancer 064003	12.4
Melanoma Margin (Lung)	7.5	Normal Bladder	18.0
Normal Kidney	4.0	Bladder Cancer 1023	11.7
Kidney Ca, Nuclear grade 2 (OD04338)	39.2	Bladder Cancer A302173	5.6
Kidney Margin (OD04338)	6.4	Normal Stomach	39.5
Kidney Ca Nuclear grade 1/2 (OD04339)	39.0	Gastric Cancer 9060397	24.5
Kidney Margin (OD04339)	3.8	Stomach Margin 9060396	28.3
Kidney Ca, Clear cell type (OD04340)	51.1	Gastric Cancer 9060395	10.0
Kidney Margin (OD04340)	16.8	Stomach Margin 9060394	29.9
Kidney Ca, Nuclear grade 3 (OD04348)	4.9	Gastric Cancer 064005	25.2

Table ABF. general oncology screening panel\_v\_2.4

Tissue Name	Rel. Exp.(%) Ag2620, Run 259737766	Tissue Name	Rel. Exp.(%) Ag2620, Run 259737766
Colon cancer 1	67.8	Bladder cancer NAT 2	0.0
Colon cancer NAT 1	17.2	Bladder cancer NAT 3	1.8

Colon cancer 2	48.6	Bladder cancer NAT 4	2.8
Colon cancer NAT 2	5.7	Prostate adenocarcinoma 1	4.6
Colon cancer 3	49.0	Prostate adenocarcinoma 2	3.4
Colon cancer NAT 3	27.5	Prostate adenocarcinoma 3	5.4
Colon malignant cancer 4	95.3	Prostate adenocarcinoma 4	93.3
Colon normal adjacent tissue 4	5.8	Prostate cancer NAT 5	4.1
Lung cancer 1	14.7	Prostate adenocarcinoma 6	0.8
Lung NAT 1	0.7	Prostate adenocarcinoma 7	3.0
Lung cancer 2	100.0	Prostate adenocarcinoma 8	1.0
Lung NAT 2	3.1	Prostate adenocarcinoma 9	5.4
Squamous cell carcinoma 3	18.7	Prostate cancer NAT 10	1.8
Lung NAT 3	1.8	Kidney cancer 1	13.6
metastatic melanoma 1	5.6	Kidney NAT 1	8.0
Melanoma 2	11.8	Kidney cancer 2	24.5
Melanoma 3	5.8	Kidney NAT 2	13.9
metastatic melanoma 4	12.2	Kidney cancer 3	38.7
metastatic melanoma 5	17.1	Kidney NAT 3	8.1
Bladder cancer 1	0.6	Kidney cancer 4	26.6
Bladder cancer NAT 1	0.0	Kidney NAT 4	15.0
Bladder cancer 2	10.0		

**General\_screening\_panel\_v1.5 Summary:** Ag2620 Highest expression of this gene is seen in a prostate cancer cell line (CT=25.9). In addition, high to moderate levels of expression are seen in all the clusters of cancer cell line samples on this panel, including brain, colon, gastric, pancreatic, renal, lung, breast, ovarian, and melanoma cancer cell

5 lines. This expression profile suggests a role for this gene product in cell survival and proliferation. Modulation of this gene product may be useful in the treatment of cancer...

This gene encodes an ephrin receptor A2-like protein (EphA2) which is activated by phosphorylation both in the tumor itself and the endothelial cells associated with the tumor. This activation is especially prominent in tumor types that are highly vascularized like colon, kidney and ovarian cancers. It appears that without the proper ligand, this  
5 overexpression and activation leads to cell transformation and the promotion of tumor-related angiogenesis which affect the overall balance between survival/apoptotic stimuli. Modifications in the signaling emanating from this receptor will impact that balance resulting either in increased survival (stimulation of angiogenesis) or increased apoptosis (inhibition of tumorigenesis both directly against tumor cells and indirectly against  
10 endothelial cells. Therefore, therapeutic targeting of this gene product with a human monoclonal antibody will affect the overall balance between survival/apoptotic stimuli in cell expressing it, preferably endothelial, tumor and neuronal cells and will therefore affect the outcome of diseases where these stimuli are involved in the pathogenesis, tumors, preferably colon, kidney and ovarian cancer, pathogenic angiogenesis, preferably wound  
15 healing, neurodegenerative diseases.

Among tissues with metabolic function, this gene is expressed at moderate to low levels in adipose, adrenal gland, pancreas, thyroid, and adult and fetal skeletal muscle, heart, and liver. This widespread expression among these tissues suggests that this gene product may play a role in normal neuroendocrine and metabolic function and that dysregulated  
20 expression of this gene may contribute to neuroendocrine disorders or metabolic diseases, such as obesity and diabetes.

This gene is also expressed at low but significant levels in the CNS, including the hippocampus, thalamus, substantia nigra, amygdala, and cerebellum. Therefore, therapeutic modulation of the expression or function of this gene may be useful in the treatment of  
25 neurologic disorders, such as Alzheimer's disease, Parkinson's disease, schizophrenia, multiple sclerosis, stroke and epilepsy.

**Oncology\_cell\_line\_screening\_panel\_v3.1 Summary:** Ag2620 Highest expression is seen in a pancreatic cancer cell line (CT=27.8). Moderate levels of expression are also seen in many of the cell lines on this panel. Please see Panel 1.5 for discussion of utility of this  
30 gene in the treatment of cancer.

**Panel 1.3D Summary:** Ag2620. Highest expression of this gene is seen in an ovarian cancer cell line (CT=29.3). In addition, moderate to low levels of expression are seen in many of the clusters of cancer cell line samples on this panel, including brain, colon, gastric, pancreatic, renal, lung, breast, ovarian, and melanoma cancer cell lines. This expression profile suggests a role for this gene product in cell survival and proliferation. Modulation of this gene product may be useful in the treatment of cancer.

Among tissues with metabolic function, this gene is expressed at low levels in adipose, pancreas, and fetal skeletal muscle, heart, and liver. This widespread expression among these tissues suggests that this gene product may play a role in normal neuroendocrine and metabolic function and that dysregulated expression of this gene may contribute to neuroendocrine disorders or metabolic diseases, such as obesity and diabetes.

In addition, this gene is expressed at much higher levels in fetal heart tissue (CT=32) when compared to expression in the adult counterpart (CT=35). Thus, expression of this gene may be used to differentiate between the fetal and adult source of this tissue.

**Panel 2.2 Summary:** Ag2620. Highest expression is seen in a sample of normal kidney (CT=31). In addition, this gene appears to be more highly expressed in kidney cancer than in the corresponding normal adjacent tissue. Thus, expression of this gene could be used as a marker of this cancer. Furthermore, therapeutic modulation of the expression or function of this gene product may be useful in the treatment of kidney cancer.

**general oncology screening panel\_v\_2.4 Summary:** Ag2620. Highest expression is seen in a sample of lung cancer (CT=29.5). In addition, this gene appears to be more highly expressed in colon and kidney cancers than in the corresponding normal adjacent tissue. Thus, expression of this gene could be used as a marker of these cancers. Furthermore, therapeutic modulation of the expression or function of this gene product may be useful in the treatment of colon and kidney cancer.

#### **AC. CG157505-01: kinesin 16A.**

Expression of gene CG157505-01 was assessed using the primer-probe set Ag5721, described in Table ACA. Results of the RTQ-PCR runs are shown in Tables ACB, ACC, and ACD.

Table ACA. Probe Name Ag5721

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5' - ctgaaggagccaatatcaacaa - 3'	22	809	555
Probe	TET-5' - tcccttgtagactctaggaattgtcatctcc - 3' - TAMRA	30	832	556
Reverse	5' - gctgaaaacttgggagttctg - 3'	21	871	557

Table ACB. CNS\_neurodegeneration\_v1.0.

Tissue Name	Rel. Exp.(%) Ag5721, Run 247018773	Tissue Name	Rel. Exp.(%) Ag5721, Run 247018773
AD 1 Hippo	18.0	Control (Path) 3 Temporal Ctx	4.9
AD 2 Hippo	16.8	Control (Path) 4 Temporal Ctx	20.4
AD 3 Hippo	10.1	AD 1 Occipital Ctx	24.8
AD 4 Hippo	7.0	AD 2 Occipital Ctx (Missing)	0.0
AD 5 hippo	87.7	AD 3 Occipital Ctx	9.6
AD 6 Hippo	27.0	AD 4 Occipital Ctx	21.9
Control 2 Hippo	21.0	AD 5 Occipital Ctx	25.5
Control 4 Hippo	11.7	AD 6 Occipital Ctx	24.8
Control (Path) 3 Hippo	5.8	Control 1 Occipital Ctx	5.1
AD 1 Temporal Ctx	40.9	Control 2 Occipital Ctx	43.2
AD 2 Temporal Ctx	25.5	Control 3 Occipital Ctx	26.1
AD 3 Temporal Ctx	5.7	Control 4 Occipital Ctx	10.3
AD 4 Temporal Ctx	24.3	Control (Path) 1 Occipital Ctx	72.2
AD 5 Inf Temporal Ctx	<b>100.0</b>	Control (Path) 2 Occipital Ctx	13.9
AD 5 Sup Temporal Ctx	52.5	Control (Path) 3 Occipital Ctx	3.5
AD 6 Inf Temporal Ctx	72.7	Control (Path) 4 Occipital Ctx	23.7

AD 6 Sup Temporal Ctx	44.4	Control 1 Parietal Ctx	8.1
Control 1 Temporal Ctx	9.0	Control 2 Parietal Ctx	65.5
Control 2 Temporal Ctx	17.6	Control 3 Parietal Ctx	18.0
Control 3 Temporal Ctx	16.8	Control (Path) 1 Parietal Ctx	34.9
Control 4 Temporal Ctx	11.7	Control (Path) 2 Parietal Ctx	26.8
Control (Path) 1 Temporal Ctx	36.1	Control (Path) 3 Parietal Ctx	2.1
Control (Path) 2 Temporal Ctx	27.0	Control (Path) 4 Parietal Ctx	39.2

Table ACC. General\_screening\_panel\_v1.5

Tissue Name	Rel. Exp.(%) Ag5721, Run 245454345	Tissue Name	Rel. Exp.(%) Ag5721, Run 245454345
Adipose	11.0	Renal ca. TK-10	18.9
Melanoma* Hs688(A).T	5.4	Bladder	6.0
Melanoma* Hs688(B).T	2.0	Gastric ca. (liver met.) NCI-N87	1.6
Melanoma* M14	13.2	Gastric ca. KATO III	0.5
Melanoma* LOXIMVI	7.6	Colon ca. SW-948	0.5
Melanoma* SK- MEL-5	4.6	Colon ca. SW480	8.3
Squamous cell carcinoma SCC-4	1.0	Colon ca.* (SW480 met) SW620	6.5
Testis Pool	28.3	Colon ca. HT29	0.1
Prostate ca.* (bone met) PC-3	6.4	Colon ca. HCT-116	16.3
Prostate Pool	10.6	Colon ca. CaCo-2	1.2
Placenta	9.7	Colon cancer tissue	5.5
Uterus Pool	48.0	Colon ca. SW1116	1.7
Ovarian ca. OVCAR-3	3.6	Colon ca. Colo-205	0.0
Ovarian ca. SK- OV-3	19.1	Colon ca. SW-48	0.0
Ovarian ca. OVCAR-4	1.4	Colon Pool	43.5

Ovarian ca. OVCAR-5	6.1	Small Intestine Pool	32.5
Ovarian ca. IGROV-1	5.4	Stomach Pool	19.2
Ovarian ca. OVCAR-8	7.1	Bone Marrow Pool	16.6
Ovary	29.5	Fetal Heart	38.4
Breast ca. MCF-7	1.0	Heart Pool	15.7
Breast ca. MDA-MB-231	15.2	Lymph Node Pool	35.4
Breast ca. BT 549	28.9	Fetal Skeletal Muscle	24.0
Breast ca. T47D	0.3	Skeletal Muscle Pool	13.7
Breast ca. MDA-N	3.2	Spleen Pool	16.4
Breast Pool	38.2	Thymus Pool	31.6
Trachea	21.9	CNS cancer (glio/astro) U87-MG	17.7
Lung	8.4	CNS cancer (glio/astro) U-118-MG	16.6
Fetal Lung	100.0	CNS cancer (neuro;met) SK-N-AS	18.9
Lung ca. NCI-N417	2.9	CNS cancer (astro) SF-539	15.9
Lung ca. LX-1	5.2	CNS cancer (astro) SNB-75	24.8
Lung ca. NCI-H146	5.5	CNS cancer (glio) SNB-19	6.3
Lung ca. SHP-77	8.8	CNS cancer (glio) SF-295	19.6
Lung ca. A549	7.2	Brain (Amygdala) Pool	11.0
Lung ca. NCI-H526	1.1	Brain (cerebellum)	31.2
Lung ca. NCI-H23	15.0	Brain (fetal)	28.1
Lung ca. NCI-H460	4.0	Brain (Hippocampus) Pool	6.6
Lung ca. HOP-62	12.2	Cerebral Cortex Pool	10.5
Lung ca. NCI-H522	20.9	Brain (Substantia nigra) Pool	10.3
Liver	0.3	Brain (Thalamus) Pool	15.5
Fetal Liver	3.3	Brain (whole)	7.7
Liver ca. HepG2	13.0	Spinal Cord Pool	13.5
Kidney Pool	71.2	Adrenal Gland	6.2
Fetal Kidney	19.8	Pituitary gland Pool	1.2
Renal ca. 786-0	11.1	Salivary Gland	2.3
Renal ca. A498	3.1	Thyroid (female)	2.0

Renal ca. ACHN	13.7	Pancreatic ca. CAPAN2	0.2
Renal ca. UO-31	5.6	Pancreas Pool	26.1

Table ACD. Panel 4.1D

Tissue Name	Rel. Exp.(%) Ag5721, Run 246509239	Tissue Name	Rel. Exp.(%) Ag5721, Run 246509239
Secondary Th1 act	36.3	HUVEC IL-1beta	13.1
Secondary Th2 act	22.8	HUVEC IFN gamma	26.2
Secondary Tr1 act	5.3	HUVEC TNF alpha + IFN gamma	0.5
Secondary Th1 rest	2.6	HUVEC TNF alpha + IL4	2.7
Secondary Th2 rest	0.0	HUVEC IL-11	14.9
Secondary Tr1 rest	2.1	Lung Microvascular EC none	40.3
Primary Th1 act	0.0	Lung Microvascular EC TNFalpha + IL-1beta	14.6
Primary Th2 act	17.7	Microvascular Dermal EC none	4.9
Primary Tr1 act	11.9	Microvascular Dermal EC TNFalpha + IL-1beta	4.8
Primary Th1 rest	0.4	Bronchial epithelium TNFalpha + IL1beta	2.3
Primary Th2 rest	5.1	Small airway epithelium none	4.8
Primary Tr1 rest	1.1	Small airway epithelium TNFalpha + IL-1beta	4.2
CD45RA CD4 lymphocyte act	17.2	Coronary artery SMC rest	3.0
CD45RO CD4 lymphocyte act	23.8	Coronary artery SMC TNFalpha + IL-1beta	4.6
CD8 lymphocyte act	2.5	Astrocytes rest	0.0
Secondary CD8 lymphocyte rest	14.9	Astrocytes TNFalpha + IL-1beta	0.8
Secondary CD8 lymphocyte act	1.5	KU-812 (Basophil) rest	0.8
CD4 lymphocyte none	0.7	KU-812 (Basophil) PMA/ionomycin	3.6
2ry Th1/Th2/Tr1_anti- CD95 CH11	5.8	CCD1106 (Keratinocytes) none	12.3
LAK cells rest	3.2	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	11.3

LAK cells IL-2	2.7	Liver cirrhosis	6.2
LAK cells IL-2+IL-12	0.0	NCI-H292 none	1.4
LAK cells IL-2+IFN gamma	4.9	NCI-H292 IL-4	5.8
LAK cells IL-2+ IL-18	1.3	NCI-H292 IL-9	4.8
LAK cells PMA/ionomycin	3.5	NCI-H292 IL-13	2.1
NK Cells IL-2 rest	94.6	NCI-H292 IFN gamma	0.9
Two Way MLR 3 day	4.5	HPAEC none	8.9
Two Way MLR 5 day	1.5	HPAEC TNF alpha + IL-1 beta	20.9
Two Way MLR 7 day	2.3	Lung fibroblast none	14.6
PBMC rest	1.5	Lung fibroblast TNF alpha + IL-1 beta	10.2
PBMC PWM	1.8	Lung fibroblast IL-4	1.5
PBMC PHA-L	3.6	Lung fibroblast IL-9	3.4
Ramos (B cell) none	4.7	Lung fibroblast IL-13	2.3
Ramos (B cell) ionomycin	26.4	Lung fibroblast IFN gamma	6.0
B lymphocytes PWM	4.9	Dermal fibroblast CCD1070 rest	18.7
B lymphocytes CD40L and IL-4	13.7	Dermal fibroblast CCD1070 TNF alpha	100.0
EOL-1 dbcAMP	14.7	Dermal fibroblast CCD1070 IL-1 beta	8.4
EOL-1 dbcAMP PMA/ionomycin	0.6	Dermal fibroblast IFN gamma	19.3
Dendritic cells none	8.7	Dermal fibroblast IL-4	43.5
Dendritic cells LPS	0.7	Dermal Fibroblasts rest	22.7
Dendritic cells anti-CD40	0.6	Neutrophils TNFa+LPS	0.8
Monocytes rest	0.0	Neutrophils rest	1.3
Monocytes LPS	2.0	Colon	5.1
Macrophages rest	1.5	Lung	2.6
Macrophages LPS	0.0	Thymus	12.1
HUVEC none	9.3	Kidney	11.0
HUVEC starved	13.3		

**CNS\_neurodegeneration\_v1.0 Summary:** Ag5721. This panel confirms the expression of this gene at moderate levels in the brain in an independent group of individuals. This gene is found to be upregulated in the temporal cortex of Alzheimer's disease patients. This gene encodes a putative kinesin, a microtubule-based motor protein involved in the transport of

organelles. Axonal transport of APP in neurons is mediated by binding with kinesin. (Gunewardena S, Neuron 2001 Nov. 8;32(3):389-401). Kamal et al. suggest that impaired APP transport leads to enhanced axonal generation and deposition of Abeta, resulting in disruption of neurotrophic signaling and neurodegeneration (Nature 2001 Dec 6;414(6864):643-8). Thus, therapeutic modulation of the expression or function of this gene may be useful in the treatment of neurodegenerative disorders, and specifically may decrease neuronal death and be of use in the treatment of Alzheimer's disease.

**General\_screening\_panel\_v1.5 Summary:** Ag5721 Highest expression of this gene is seen in the fetal lung (CT=27.5). In addition, this gene is expressed at much higher levels in fetal lung tissue when compared to expression in the adult counterpart (CT=31). Thus, expression of this gene may be used to differentiate between the fetal and adult source of this tissue. In addition, therapeutic modulation of the expression or function of this gene may be useful in the treatment of diseases that affect the lung, including lung cancer.

Moderate to low levels of expression are seen in all regions of the CNS examined. Please see CNS\_neurodegeneration\_v1.0 for discussion of utility of this gene in CNS disorders.

Moderate to low levels of expression are also seen in pancreas, thyroid, fetal skeletal muscle, adipose and adult and fetal liver. This widespread expression among these tissues suggests that this gene product may play a role in normal neuroendocrine and metabolic function and that dysregulated expression of this gene may contribute to neuroendocrine disorders or metabolic diseases, such as obesity and diabetes.

Low but significant levels of expression are seen in many of the cancer cell lines on this panel. Interestingly, expression appears to be overexpressed in the normal tissue samples when compared to expression in the cell lines. Thus, modulation of the expression or function of this gene may be useful in the treatment of cancer.

**Panel 4.1D Summary:** Ag5721 Highest expression of this gene is seen in TNF-alpha treated dermal fibroblasts (CT=30.2). Moderate levels of expression are also seen in resting NK cells. Low but significant levels of expression are seen in activated T cells, endothelial cells and lung and dermal fibroblasts. Thus, expression of this gene could be used as a marker of activated dermal fibroblasts and modulation of the gene product may be useful in the treatment of psoriasis.

**AD. CG157629-01: SERINE/THREONINE PROTEIN PHOSPHATASE WITH EF-HANDS-1.**

Expression of gene CG157629-01 was assessed using the primer-probe set Ag5447, described in Table ADA. Please note that CG157629-01 represents a full-length physical  
5 clone.

Table ADA. Probe Name Ag5447

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-ctggctcccaacgga-3'	15	906	558
Probe	TET-5'-tggatctcctactgaacacttaacagagcatg-3'-TAMRA	32	1002	559
Reverse	5'-acagaatatcaataatctgttcccat-3'	26	1035	560

**AI\_comprehensive panel\_v1.0 Summary:** Ag5447 Expression of this gene is low/undetectable in all samples on this panel (CTs>35). (Data not shown.)

**General\_screening\_panel\_v1.5 Summary:** Ag5447 Expression of this gene is  
10 low/undetectable in all samples on this panel (CTs>35). (Data not shown.)

**Panel 4.1D Summary:** Ag5447 Expression of this gene is low/undetectable in all samples on this panel (CTs>35). (Data not shown.)

**AE. CG157704-01: kinesin 24.**

Expression of gene CG157704-01 was assessed using the primer-probe set Ag5734,  
15 described in Table AEA. Results of the RTQ-PCR runs are shown in Tables AEB, AEC and AED.

Table AEA. Probe Name Ag5734

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-gaggtacgtcgtggagaaatta-3'	22	718	561
Probe	TET-5'-tcatgcacaagtagagtttctttgtcttc-3'-TAMRA	29	754	562
Reverse	5'-tgaggtcaactgcttctttctt-3'	22	784	563

Table AEB. CNS\_neurodegeneration\_v1.0

Tissue Name	Rel. Exp.(%) Ag5734, Run 247018774	Tissue Name	Rel. Exp.(%) Ag5734, Run 247018774
AD 1 Hippo	15.3	Control (Path) 3 Temporal Ctx	2.6
AD 2 Hippo	15.9	Control (Path) 4 Temporal Ctx	64.6
AD 3 Hippo	9.0	AD 1 Occipital Ctx	20.2
AD 4 Hippo	8.7	AD 2 Occipital Ctx (Missing)	0.0
AD 5 Hippo	68.8	AD 3 Occipital Ctx	7.7
AD 6 Hippo	57.4	AD 4 Occipital Ctx	24.5
Control 2 Hippo	29.1	AD 5 Occipital Ctx	33.0
Control 4 Hippo	24.3	AD 6 Occipital Ctx	18.4
Control (Path) 3 Hippo	20.4	Control 1 Occipital Ctx	16.4
AD 1 Temporal Ctx	17.8	Control 2 Occipital Ctx	43.8
AD 2 Temporal Ctx	36.9	Control 3 Occipital Ctx	20.6
AD 3 Temporal Ctx	13.9	Control 4 Occipital Ctx	25.2
AD 4 Temporal Ctx	24.5	Control (Path) 1 Occipital Ctx	100.0
AD 5 Inf Temporal Ctx	74.7	Control (Path) 2 Occipital Ctx	16.4
AD 5 Sup Temporal Ctx	41.8	Control (Path) 3 Occipital Ctx	0.0
AD 6 Inf Temporal Ctx	42.3	Control (Path) 4 Occipital Ctx	22.1
AD 6 Sup Temporal Ctx	66.4	Control 1 Parietal Ctx	18.3
Control 1 Temporal Ctx	20.2	Control 2 Parietal Ctx	23.3
Control 2 Temporal Ctx	33.4	Control 3 Parietal Ctx	11.7
Control 3 Temporal Ctx	15.7	Control (Path) 1 Parietal Ctx	43.5
Control 3 Temporal Ctx	3.0	Control (Path) 2 Parietal Ctx	20.3
Control (Path) 1. Temporal Ctx	50.0	Control (Path) 3 Parietal Ctx	14.0
Control (Path) 2 Temporal Ctx	39.0	Control (Path) 4 Parietal Ctx	29.1

Table AEC. General\_screening\_panel\_v1.5.

Tissue Name	Rel. Exp.(%) Ag5734, Run 245385008	Tissue Name	Rel. Exp.(%) Ag5734, Run 245385008
Adipose	0.3	Renal ca. TK-10	22.2
Melanoma* Hs688(A).T	2.7	Bladder	10.2
Melanoma* Hs688(B).T	1.4	Gastric ca. (liver met.) NCI-N87	50.0
Melanoma* M14	29.7	Gastric ca. KATO III	100.0
Melanoma* LOXIMVI	36.3	Colon ca. SW-948	6.1
Melanoma* SK- MEL-5	19.3	Colon ca. SW480	68.3
Squamous cell carcinoma SCC-4	13.2	Colon ca.* (SW480 met) SW620	44.4
Testis Pool	3.3	Colon ca. HT29	23.8
Prostate ca.* (bone met) PC-3	7.5	Colon ca. HCT-116	42.0
Prostate Pool	1.1	Colon ca. CaCo-2	19.5
Placenta	3.8	Colon cancer tissue	10.0
Uterus Pool	1.3	Colon ca. SW1116	7.4
Ovarian ca. OVCAR-3	40.1	Colon ca. Colo-205	9.4
Ovarian ca. SK- OV-3	1.3	Colon ca. SW-48	11.7
Ovarian ca. OVCAR-4	9.4	Colon Pool	0.0
Ovarian ca. OVCAR-5	31.2	Small Intestine Pool	5.0
Ovarian ca. IGROV-1	10.9	Stomach Pool	1.9
Ovarian ca. OVCAR-8	9.0	Bone Marrow Pool	1.3
Ovary	3.8	Fetal Heart	6.8
Breast ca. MCF-7	13.7	Heart Pool	2.0
Breast ca. MDA- MB-231	77.9	Lymph Node Pool	3.3
Breast ca. BT 549	89.5	Fetal Skeletal Muscle	0.0
Breast ca. T47D	15.8	Skeletal Muscle Pool	2.1
Breast ca. MDA-N	17.8	Spleen Pool	1.4
Breast Pool	2.9	Thymus Pool	16.3
Trachea	10.1	CNS cancer (glio/astro) U87-MG	47.6
Lung	1.1	CNS cancer	81.2

		(glio/astro) U-118-MG	
Fetal Lung	23.2	CNS cancer (neuro;met) SK-N-AS	26.4
Lung ca. NCI-N417	4.9	CNS cancer (astro) SF-539	26.1
Lung ca. LX-1	46.7	CNS cancer (astro) SNB-75	75.8
Lung ca. NCI-H146	27.0	CNS cancer (glio) SNB-19	8.4
Lung ca. SHP-77	31.4	CNS cancer (glio) SF-295	20.9
Lung ca. A549	44.1	Brain (Amygdala) Pool	1.4
Lung ca. NCI-H526	10.0	Brain (cerebellum)	5.7
Lung ca. NCI-H23	1.7	Brain (fetal)	11.0
Lung ca. NCI-H460	0.1	Brain (Hippocampus) Pool	2.8
Lung ca. HOP-62	3.5	Cerebral Cortex Pool	4.8
Lung ca. NCI-H522	17.3	Brain (Substantia nigra) Pool	2.9
Liver	0.1	Brain (Thalamus) Pool	4.6
Fetal Liver	28.5	Brain (whole)	4.8
Liver ca. HepG2	1.3	Spinal Cord Pool	4.0
Kidney Pool	6.0	Adrenal Gland	3.2
Fetal Kidney	19.2	Pituitary gland Pool	2.4
Renal ca. 786-0	23.3	Salivary Gland	1.1
Renal ca. A498	9.3	Thyroid (female)	3.5
Renal ca. ACHN	7.5	Pancreatic ca. CAPAN2	23.0
Renal ca. UO-31	10.2	Pancreas Pool	1.9

Table AED. Panel 4.1D

Tissue Name	Rel. Exp.(%) Ag5734, Run 246509244	Tissue Name	Rel. Exp.(%) Ag5734, Run 246509244
Secondary Th1 act	65.5	HUVEC IL-1beta	19.5
Secondary Th2 act	98.6	HUVEC IFN gamma	21.5
Secondary Tr1 act	20.9	HUVEC TNF alpha + IFN gamma	2.1
Secondary Th1 rest	0.0	HUVEC TNF alpha + IL4	1.8
Secondary Th2 rest	0.0	HUVEC IL-11	9.0
Secondary Tr1 rest	0.0	Lung Microvascular EC none	12.2

Primary Th1 act	0.4	Lung Microvascular EC TNFalpha + IL-1beta	2.7
Primary Th2 act	13.8	Microvascular Dermal EC none	0.4
Primary Tr1 act	9.5	Microvascular Dermal EC TNFalpha + IL-1beta	4.0
Primary Th1 rest	0.0	Bronchial epithelium TNFalpha + IL1beta	3.7
Primary Th2 rest	0.0	Small airway epithelium none	1.3
Primary Tr1 rest	0.0	Small airway epithelium TNFalpha + IL-1beta	4.5
CD45RA CD4 lymphocyte act	30.4	Coronary artery SMC rest	3.5
CD45RO CD4 lymphocyte act	43.2	Coronary artery SMC TNFalpha + IL-1beta	2.9
CD8 lymphocyte act	3.9	Astrocytes rest	3.4
Secondary CD8 lymphocyte rest	17.0	Astrocytes TNFalpha + IL-1beta	0.9
Secondary CD8 lymphocyte act	3.3	KU-812 (Basophil) rest	29.9
CD4 lymphocyte none	0.0	KU-812 (Basophil) PMA/ionomycin	40.9
2ry Th1/Th2/Tr1_anti- CD95 CH11	1.7	CCD1106 (Keratinocytes) none	47.0
LAK cells rest	8.1	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	20.7
LAK cells IL-2	7.1	Liver cirrhosis	2.0
LAK cells IL-2+IL-12	2.7	NCI-H292 none	26.6
LAK cells IL-2+IFN gamma	4.0	NCI-H292 IL-4	30.6
LAK cells IL-2+ IL-18	2.3	NCI-H292 IL-9	63.7
LAK cells PMA/ionomycin	15.8	NCI-H292 IL-13	29.3
NK Cells IL-2 rest	77.4	NCI-H292 IFN gamma	16.0
Two Way MLR 3 day	4.5	HPAEC none	3.5
Two Way MLR 5 day	1.6	HPAEC TNF alpha + IL- 1 beta	12.1
Two Way MLR 7 day	6.9	Lung fibroblast none	3.9
PBMC rest	0.0	Lung fibroblast TNF alpha + IL-1 beta	5.4
PBMC PWM	3.8	Lung fibroblast IL-4	1.0
PBMC PHA-L	8.8	Lung fibroblast IL-9	6.2

Ramos (B cell) none	4.9	Lung fibroblast IL-13	0.0
Ramos (B cell) ionomycin	35.4	Lung fibroblast IFN gamma	5.4
B lymphocytes PWM	24.0	Dermal fibroblast CCD1070 rest	46.7
B lymphocytes CD40L and IL-4	45.7	Dermal fibroblast CCD1070 TNF alpha	100.0
EOL-1 dbcAMP	60.7	Dermal fibroblast CCD1070 IL-1 beta	22.5
EOL-1 dbcAMP PMA/ionomycin	3.2	Dermal fibroblast IFN gamma	16.6
Dendritic cells none	6.3	Dermal fibroblast IL-4	19.9
Dendritic cells LPS	0.7	Dermal Fibroblasts rest	3.7
Dendritic cells anti-CD40	1.6	Neutrophils TNFa+LPS	1.6
Monocytes rest	1.6	Neutrophils rest	2.6
Monocytes LPS	3.7	Colon	0.7
Macrophages rest	3.8	Lung	0.6
Macrophages LPS	0.8	Thymus	12.3
HUVEC none	10.1	Kidney	6.8
HUVEC starved	36.9		

**CNS\_neurodegeneration\_v1.0 Summary:** Ag5734 This panel does not show differential expression of this gene in Alzheimer's disease. However, this profile confirms the expression of this gene at moderate levels in the brain. Please see Panel 1.5 for discussion of utility of this gene in the central nervous system.

- 5 **General\_screening\_panel\_v1.5 Summary:** Ag5734 Highest expression of this gene is seen in a gastric cancer cell line (CT=29). This gene is widely expressed in this panel, with moderate expression seen in brain, colon, gastric, lung, breast, pancreatic, renal, ovarian, and melanoma cancer cell lines. This expression profile with prominent cell line expression suggests a role for this gene product in cell survival and proliferation. Modulation of this
- 10 gene product may be useful in the treatment of cancer.

Among tissues with metabolic function, this gene is expressed at low but significant levels in pituitary, skeletal muscle, adrenal gland, pancreas, thyroid, fetal liver, and adult and fetal liver. This widespread expression among these tissues suggests that this gene product may play a role in normal neuroendocrine and metabolic function and that dysregulated

expression of this gene may contribute to neuroendocrine disorders or metabolic diseases, such as obesity and diabetes.

This gene is also expressed at low but significant levels in the CNS, including the hippocampus, thalamus, substantia nigra, amygdala, cerebellum and cerebral cortex.

- 5 Therefore, therapeutic modulation of the expression or function of this gene may be useful in the treatment of neurologic disorders, such as Alzheimer's disease, Parkinson's disease, schizophrenia, multiple sclerosis, stroke and epilepsy.

- Panel 4.1D Summary:** Ag5734 Highest expression is seen in TNF- $\alpha$  treated dermal fibroblasts. Low but significant expression is seen in activated T cells, resting NK cells, eosinophils, activated B cells, HUVECs, basophils and NCI-H292 goblet cells. This expression suggests that this gene product may be involved in autoinflammatory processes. Thus, expression of this gene could be used as a marker of activated dermal fibroblasts. Modulation of the expression or function of this gene may be useful in the treatment of RA, OA, lupus, asthma, allergy, emphysema, and psoriasis.

- 15 **AF. CG158218-01: kinesin 6.**

Expression of gene CG158218-01 was assessed using the primer-probe set Ag5797, described in Table AFA. Results of the RTQ-PCR runs are shown in Tables AFB and AFC.

Table AFA. Probe Name Ag5797

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-agttacaaaaggacagcagcaa-3'	22	621	564
Probe	TET-5'-ccacattcattgtagattccaaatagga-3'-TAMRA	29	662	565
Reverse	5'-ttcatgtcttgatccaaaaga-3'	22	697	566

Table AFB. CNS\_neurodegeneration\_v1.0

Tissue Name	Rel. Exp.(%) Ag5797, Run 247179625	Tissue Name	Rel. Exp.(%) Ag5797, Run 247179625
AD 1 Hippo	15.9	Control (Path) 3 Temporal Ctx	4.8
AD 2 Hippo	32.1	Control (Path) 4 Temporal Ctx	22.5
AD 3 Hippo	6.8	AD 1 Occipital Ctx	12.8

AD 4 Hippo	9.5	AD 2 Occipital Ctx (Missing)	0.0
AD 5 Hippo	27.4	AD 3 Occipital Ctx	7.3
AD 6 Hippo	33.9	AD 4 Occipital Ctx	16.8
Control 2 Hippo	31.0	AD 5 Occipital Ctx	26.2
Control 4 Hippo	25.2	AD 6 Occipital Ctx	10.7
Control (Path) 3 Hippo	7.9	Control 1 Occipital Ctx	3.1
AD 1 Temporal Ctx	80.7	Control 2 Occipital Ctx	29.5
AD 2 Temporal Ctx	33.2	Control 3 Occipital Ctx	15.9
AD 3 Temporal Ctx	9.3	Control 4 Occipital Ctx	13.6
AD 4 Temporal Ctx	24.0	Control (Path) 1 Occipital Ctx	85.9
AD 5 Inf Temporal Ctx	100.0	Control (Path) 2 Occipital Ctx	11.0
AD 5 Sup Temporal Ctx	51.1	Control (Path) 3 Occipital Ctx	3.5
AD 6 Inf Temporal Ctx	35.4	Control (Path) 4 Occipital Ctx	12.7
AD 6 Sup Temporal Ctx	29.1	Control 1 Parietal Ctx	15.3
Control 1 Temporal Ctx	7.0	Control 2 Parietal Ctx	51.4
Control 2 Temporal Ctx	22.5	Control 3 Parietal Ctx	8.2
Control 3 Temporal Ctx	20.6	Control (Path) 1 Parietal Ctx	65.1
Control 3 Temporal Ctx	5.6	Control (Path) 2 Parietal Ctx	25.3
Control (Path) 1 Temporal Ctx	48.0	Control (Path) 3 Parietal Ctx	2.4
Control (Path) 2 Temporal Ctx	29.5	Control (Path) 4 Parietal Ctx	30.4

Table AFC. General\_screening\_panel\_v1.5

Tissue Name	Rel. Exp.(%) Ag5797, Run 245382863	Tissue Name	Rel. Exp.(%) Ag5797, Run 245382863
Adipose	0.3	Renal ca. TK-10	0.1
Melanoma* Hs688(A).T	0.1	Bladder	0.6

Melanoma* Hs688(B).T	0.0	Gastric ca. (liver met.) NCI-N87	0.0
Melanoma* M14	0.7	Gastric ca. KATO III	0.0
Melanoma* LOXIMVI	0.0	Colon ca. SW-948	0.0
Melanoma* SK- MEL-5	0.6	Colon ca. SW480	4.2
Squamous cell carcinoma SCC-4	0.0	Colon ca.* (SW480 met) SW620	10.8
Testis Pool	9.9	Colon ca. HT29	0.0
Prostate ca.* (bone met) PC-3	0.0	Colon ca. HCT-116	0.0
Prostate Pool	0.6	Colon ca. CaCo-2	0.2
Placenta	0.1	Colon cancer tissue	0.0
Uterus Pool	0.2	Colon ca. SW1116	0.0
Ovarian ca. OVCAR-3	1.5	Colon ca. Colo-205	0.0
Ovarian ca. SK- OV-3	2.0	Colon ca. SW-48	0.1
Ovarian ca. OVCAR-4	0.0	Colon Pool	0.4
Ovarian ca. OVCAR-5	1.2	Small Intestine Pool	1.2
Ovarian ca. IGROV-1	0.1	Stomach Pool	0.6
Ovarian ca. OVCAR-8	0.0	Bone Marrow Pool	0.2
Ovary	1.4	Fetal Heart	0.0
Breast ca. MCF-7	0.3	Heart Pool	0.3
Breast ca. MDA- MB-231	0.0	Lymph Node Pool	1.0
Breast ca. BT 549	0.0	Fetal Skeletal Muscle	0.2
Breast ca. T47D	0.5	Skeletal Muscle Pool	0.1
Breast ca. MDA-N	0.2	Spleen Pool	0.1
Breast Pool	1.3	Thymus Pool	1.4
Trachea	4.2	CNS cancer (glio/astro) U87-MG	2.3
Lung	0.1	CNS cancer (glio/astro) U-118-MG	0.0
Fetal Lung	11.7	CNS cancer (neuro;met) SK-N-AS	0.4
Lung ca. NCI-N417	1.4	CNS cancer (astro) SF- 539	0.0
Lung ca. LX-1	7.4	CNS cancer (astro)	0.7

		SNB-75	
Lung ca. NCI-H146	0.0	CNS cancer (glio) SNB-19	0.4
Lung ca. SHP-77	1.7	CNS cancer (glio) SF-295	0.7
Lung ca. A549	0.0	Brain (Amygdala) Pool	7.1
Lung ca. NCI-H526	0.2	Brain (cerebellum)	2.7
Lung ca. NCI-H23	0.3	Brain (fetal)	2.1
Lung ca. NCI-H460	0.0	Brain (Hippocampus) Pool	3.7
Lung ca. HOP-62	0.2	Cerebral Cortex Pool	6.3
Lung ca. NCI-H522	0.1	Brain (Substantia nigra) Pool	9.7
Liver	0.0	Brain (Thalamus) Pool	4.0
Fetal Liver	100.0	Brain (whole)	2.8
Liver ca. HepG2	0.0	Spinal Cord Pool	11.4
Kidney Pool	0.7	Adrenal Gland	0.3
Fetal Kidney	4.7	Pituitary gland Pool	1.7
Renal ca. 786-0	0.1	Salivary Gland	0.0
Renal ca. A498	0.1	Thyroid (female)	0.7
Renal ca. ACHN	0.1	Pancreatic ca. CAPAN2	0.3
Renal ca. UO-31	0.4	Pancreas Pool	0.8

**CNS\_neurodegeneration\_v1.0 Summary:** Ag5797 This panel does not show differential expression of this gene in Alzheimer's disease. However, this profile confirms the expression of this gene at moderate levels in the brain. Please see Panel 1.5 for discussion of utility of this gene in the central nervous system.

- 5 **General\_screening\_panel\_v1.5 Summary:** Ag5797 Highest expression of this gene is seen in the fetal liver. Interestingly, this gene is expressed at much higher levels in fetal (CT = 29) when compared to adult liver tissue (CT = 40). This observation suggests that expression of this gene can be used to distinguish fetal from adult liver. In addition, the relative overexpression of this gene in fetal liver suggests that the protein product may
- 10 enhance liver growth or development in the fetus and thus may also act in a regenerative capacity in the adult. Therefore, therapeutic modulation of the protein encoded by this gene could be useful in treatment of liver related diseases.

- This gene is also expressed at low levels in the CNS, including the hippocampus, thalamus, substantia nigra, amygdala, cerebellum and cerebral cortex. Therefore, therapeutic modulation of the expression or function of this gene may be useful in the treatment of neurological disorders, such as Alzheimer's disease, Parkinson's disease, schizophrenia, multiple sclerosis, stroke and epilepsy.

**Panel 4.1D Summary:** Ag5797 Expression of this gene is low/undetectable in all samples on this panel (CTs>35). (Data not shown.)

**AG. CG158583-01 and CG158583-04: SYNAPTIC VESICLE AMINE TRANSPORTER.**

- 10 Expression of gene CG158583-01 and CG158583-04 was assessed using the primer-probe set Ag7590, described in Table AGA. Results of the RTQ-PCR runs are shown in Table AGB. Please note that CG158583-04 represents a full-length physical clone.

Table AGA. Probe Name Ag7590

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5' -aactcctgacctcaggtgatc-3'	21	167	567
Probe	TET-5' -tcctggaattacagtcctccatcatcc-3' - TAMRA	26	210	568
Reverse	5' -ctcatgcttaatgctgtacagataact-3'	27	238	569

Table AGB. Panel 5 Islet

Tissue Name	Rel. Exp.(%) Ag7590, Run 310258790	Tissue Name	Rel. Exp.(%) Ag7590, Run 310258790
97457_Patient-02go_adipose	0.0	94709_Donor 2 AM - A_adipose	0.0
97476_Patient-07sk_skeletal muscle	0.0	94710_Donor 2 AM - B_adipose	0.0
97477_Patient-07ut_uterus	0.0	94711_Donor 2 AM - C_adipose	0.0
97478_Patient-07pl_placenta	0.0	94712_Donor 2 AD - A_adipose	0.0
99167_Bayer Patient 1	100.0	94713_Donor 2 AD - B_adipose	0.0
97482_Patient-08ut_uterus	12.2	94714_Donor 2 AD - C_adipose	0.0

97483_Patient-08pl_placenta	0.0	94742_Donor 3 U - A_Mesenchymal Stem Cells	0.0
97486_Patient-09sk_skeletal muscle	10.2	94743_Donor 3 U - B_Mesenchymal Stem Cells	0.0
97487_Patient-09ut_uterus	0.0	94730_Donor 3 AM - A_adipose	0.0
97488_Patient-09pl_placenta	0.0	94731_Donor 3 AM - B_adipose	0.0
97492_Patient-10ut_uterus	21.6	94732_Donor 3 AM - C_adipose	0.0
97493_Patient-10pl_placenta	0.0	94733_Donor 3 AD - A_adipose	0.0
97495_Patient-11go_adipose	0.0	94734_Donor 3 AD - B_adipose	0.0
97496_Patient-11sk_skeletal muscle	27.2	94735_Donor 3 AD - C_adipose	0.0
97497_Patient-11ut_uterus	0.0	77138_Liver_HepG2untreated	0.0
97498_Patient-11pl_placenta	0.0	73556_Heart_Cardiac stromal cells (primary)	0.0
97500_Patient-12go_adipose	32.3	81735_Small Intestine	26.6
97501_Patient-12sk_skeletal muscle	0.0	72409_Kidney Proximal Convoluted Tubule	0.0
97502_Patient-12ut_uterus	13.3	82685_Small intestine Duodenum	14.8
97503_Patient-12pl_placenta	0.0	90650_Adrenal_Adrenocortical adenoma	0.0
94721_Donor 2 U - A_Mesenchymal Stem Cells	0.0	72410_Kidney_HRCE	0.0
94722_Donor 2 U - B_Mesenchymal Stem Cells	0.0	72411_Kidney_HRE	0.0
94723_Donor 2 U - C_Mesenchymal Stem Cells	0.0	73139_Uterus_Uterine smooth muscle cells	0.0

**Panel 5 Islet Summary:** Ag7590 Expression of this gene is restricted to a sample of pancreatic islet cells (CT=34.5). Thus, expression of this gene could be used to differentiate between this sample and other samples on this panel and as a marker of islet cells.

Furthermore, therapeutic modulation of the expression or function of this gene may be

5 useful in the treatment of diabetes.

**AH. CG159084-01: Glutamate Decarboxylase like.**

Expression of gene CG159084-01 was assessed using the primer-probe sets Ag5799 and Ag5799, described in Tables AHA and AHB.

Table AHA. Probe Name Ag5799

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5' -agagatcaagaactccgaaagg-3'	22	1399	570
Probe	TET-5' -tgctttccatcatcatctgtgcttta-3' - TAMRA	26	1434	571
Reverse	5' -ggctggtagcttatcatgattg-3'	22	1460	572

5. Table AHB. Probe Name Ag5799

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5' -agagatcaagaactccgaaagg-3'	22	1399	573
Probe	TET-5' -tgctttccatcatcatctgtgcttta-3' - TAMRA	26	1434	574
Reverse	5' -ggctggtagcttatcatgattg-3'	22	1460	575

**CNS\_neurodegeneration\_v1.0 Summary:** Ag5799 Expression of this gene is low/undetectable in all samples on this panel (CTs>35). (Data not shown.)

**General\_screening\_panel\_v1.5 Summary:** Ag5799 Expression of this gene is low/undetectable in all samples on this panel (CTs>35). (Data not shown.)

- 10 **General\_screening\_panel\_v1.6 Summary:** Ag5799 Expression of this gene is low/undetectable in all samples on this panel (CTs>35). (Data not shown.)

**Panel 4.1D Summary:** Ag5799 Expression of this gene is low/undetectable in all samples on this panel (CTs>35). (Data not shown.)

- 15 **Panel 5 Islet Summary:** Ag5799 Expression of this gene is low/undetectable in all samples on this panel (CTs>35). (Data not shown.)

**Panel CNS\_1.1 Summary:** Ag5799 Expression of this gene is low/undetectable in all samples on this panel (CTs>35). (Data not shown.)

**AI. CG159130-01: HYPERPOLARIZATION-ACTIVATED CYCLIC NUCLEOTIDE-GATED CHANNEL 1.**

Expression of gene CG159130-01 was assessed using the primer-probe set Ag7494, described in Table AIA. Results of the RTQ-PCR runs are shown in Table AIB.

5 Table AIA. Probe Name Ag7494

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-ttcatacgcactcttcaaagcta-3'	23	1095	576
Probe	TET-5'-cccagtcagcatgtctgacctctgga-3'-TAMRA	26	1155	577
Reverse	5'-cgacgatcatgctcagcat-3'	19	1186	578

Table AIB. CNS\_neurodegeneration\_v1.0

Tissue Name	Rel. Exp.(%) Ag7494, Run 308752180	Tissue Name	Rel. Exp.(%) Ag7494, Run 308752180
AD 1 Hippo	2.1	Control (Path) 3 Temporal Ctx	0.8
AD 2 Hippo	7.9	Control (Path) 4 Temporal Ctx	13.7
AD 3 Hippo	2.2	AD 1 Occipital Ctx	6.8
AD 4 Hippo	2.0	AD 2 Occipital Ctx (Missing)	0.0
AD 5 hippo	100.0	AD 3 Occipital Ctx	1.6
AD 6 Hippo	17.6	AD 4 Occipital Ctx	8.8
Control 2 Hippo	21.6	AD 5 Occipital Ctx	12.2
Control 4 Hippo	1.1	AD 6 Occipital Ctx	57.4
Control (Path) 3 Hippo	0.6	Control 1 Occipital Ctx	0.5
AD 1 Temporal Ctx	3.0	Control 2 Occipital Ctx	70.2
AD 2 Temporal Ctx	9.1	Control 3 Occipital Ctx	7.4
AD 3 Temporal Ctx	1.0	Control 4 Occipital Ctx	1.1
AD 4 Temporal Ctx	5.1	Control (Path) 1 Occipital Ctx	62.9

AD 5 Inf Temporal Ctx	69.3	Control (Path) 2 Occipital Ctx	3.8
AD 5 Sup Temporal Ctx	15.0	Control (Path) 3 Occipital Ctx	0.6
AD 6 Inf Temporal Ctx	14.6	Control (Path) 4 Occipital Ctx	7.2
AD 6 Sup Temporal Ctx	19.8	Control 1 Parietal Ctx	0.9
Control 1 Temporal Ctx	0.6	Control 2 Parietal Ctx	16.4
Control 2 Temporal Ctx	34.9	Control 3 Parietal Ctx	11.5
Control 3 Temporal Ctx	6.2	Control (Path) 1 Parietal Ctx	66.0
Control 4 Temporal Ctx	1.8	Control (Path) 2 Parietal Ctx	11.7
Control (Path) 1 Temporal Ctx	43.5	Control (Path) 3 Parietal Ctx	0.9
Control (Path) 2 Temporal Ctx	19.6	Control (Path) 4 Parietal Ctx	31.9

**CNS\_neurodegeneration\_v1.0 Summary:** Ag7494 This panel does not show differential expression of this gene in Alzheimer's disease. However, this profile confirms the expression of this gene at high to moderate levels in the brain. Therefore, therapeutic modulation of the expression or function of this gene may be useful in the treatment of neurological disorders, such as Alzheimer's disease, Parkinson's disease, schizophrenia, multiple sclerosis, stroke and epilepsy.

**AJ. CG159178-01: Carbonic anhydrase VI precursor.**

Expression of gene CG159178-01 was assessed using the primer-probe set Ag4880, described in Table AJA. Results of the RTQ-PCR runs are shown in Tables AJB, AJC and AJD.

Table AJA. Probe Name Ag4880

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-ttcggttgaggtgaagaattacc-3'	22	319	579
Probe	TET-5'-cagcaacttcatttctcatctggcca-3'-TAMRA	26	357	580
Reverse	5'-gttctttgtcctgggtacttga-3'	22	386	581

Table AJB. General\_screening\_panel\_v1.5

Tissue Name	Rel. Exp.(%) Ag4880, Run 228806989	Tissue Name	Rel. Exp.(%) Ag4880, Run 228806989
Adipose	0.0	Renal ca. TK-10	0.0
Melanoma* Hs688(A).T	0.0	Bladder	0.0
Melanoma* Hs688(B).T	0.0	Gastric ca. (liver met.) NCI-N87	0.0
Melanoma* M14	0.0	Gastric ca. KATO III	0.0
Melanoma* LOXIMVI	0.0	Colon ca. SW-948	0.0
Melanoma* SK- MEL-5	0.0	Colon ca. SW480	0.0
Squamous cell carcinoma SCC-4	0.0	Colon ca.* (SW480 met) SW620	0.0
Testis Pool	0.0	Colon ca. HT29	0.0
Prostate ca.* (bone met) PC-3	0.0	Colon ca. HCT-116	0.0
Prostate Pool	0.0	Colon ca. CaCo-2	0.0
Placenta	0.0	Colon cancer tissue	0.0
Uterus Pool	0.0	Colon ca. SW1116	0.0
Ovarian ca. OVCAR-3	0.0	Colon ca. Colo-205	0.0
Ovarian ca. SK- OV-3	0.0	Colon ca. SW-48	0.0
Ovarian ca. OVCAR-4	0.0	Colon Pool	0.0
Ovarian ca. OVCAR-5	0.0	Small Intestine Pool	0.0
Ovarian ca. IGROV-1	0.0	Stomach Pool	0.0
Ovarian ca. OVCAR-8	0.0	Bone Marrow Pool	0.0
Ovary	0.0	Fetal Heart	0.0
Breast ca. MCF-7	0.0	Heart Pool	0.0
Breast ca. MDA- MB-231	0.0	Lymph Node Pool	0.0
Breast ca. BT 549	0.0	Fetal Skeletal Muscle	0.0
Breast ca. T47D	0.0	Skeletal Muscle Pool	0.0
Breast ca. MDA-N	0.0	Spleen Pool	0.0
Breast Pool	0.0	Thymus Pool	0.0
Trachea	0.1	CNS cancer	0.0

		(glio/astro) U87-MG	
Lung	0.0	CNS cancer (glio/astro) U-118-MG	0.0
Fetal Lung	0.0	CNS cancer (neuro;met) SK-N-AS	0.0
Lung ca. NCI-N417	0.0	CNS cancer (astro) SF-539	0.0
Lung ca. LX-1	1.4	CNS cancer (astro) SNB-75	0.0
Lung ca. NCI-H146	0.0	CNS cancer (glio) SNB-19	0.0
Lung ca. SHP-77	0.3	CNS cancer (glio) SF-295	0.0
Lung ca. A549	0.0	Brain (Amygdala) Pool	0.0
Lung ca. NCI-H526	0.0	Brain (cerebellum)	0.0
Lung ca. NCI-H23	0.0	Brain (fetal)	0.0
Lung ca. NCI-H460	0.0	Brain (Hippocampus) Pool	0.0
Lung ca. HOP-62	0.0	Cerebral Cortex Pool	0.0
Lung ca. NCI-H522	0.0	Brain (Substantia nigra) Pool	0.0
Liver	0.0	Brain (Thalamus) Pool	0.0
Fetal Liver	0.0	Brain (whole)	0.0
Liver ca. HepG2	0.0	Spinal Cord Pool	0.0
Kidney Pool	0.0	Adrenal Gland	0.0
Fetal Kidney	0.0	Pituitary gland Pool	0.0
Renal ca. 786-0	0.0	Salivary Gland	100.0
Renal ca. A498	0.0	Thyroid (female)	0.0
Renal ca. ACHN	0.0	Pancreatic ca. CAPAN2	0.0
Renal ca. UO-31	0.0	Pancreas Pool	0.0

Table AJC. Panel 4.1D.

Tissue Name	Rel. Exp.(%) Ag4880, Run 223350178	Tissue Name	Rel. Exp.(%) Ag4880, Run 223350178
Secondary Th1 act	100.0	HUVEC IL-1beta	0.0
Secondary Th2 act	0.0	HUVEC IFN gamma	0.0
Secondary Tr1 act	7.2	HUVEC TNF alpha + IFN gamma	0.0
Secondary Th1 rest	11.3	HUVEC TNF alpha + IL4	0.0
Secondary Th2 rest	7.0	HUVEC IL-11	0.0

Secondary Tr1 rest	8.8	Lung Microvascular EC none	0.0
Primary Th1 act	5.4	Lung Microvascular EC TNFalpha + IL-1beta	0.0
Primary Th2 act	0.0	Microvascular Dermal EC none	0.0
Primary Tr1 act	43.2	Microvascular Dermal EC TNFalpha + IL-1beta	0.0
Primary Th1 rest	29.5	Bronchial epithelium TNFalpha + IL1beta	0.0
Primary Th2 rest	6.7	Small airway epithelium none	0.0
Primary Tr1 rest	10.4	Small airway epithelium TNFalpha + IL-1beta	0.0
CD45RA CD4 lymphocyte act	19.2	Coronary artery SMC rest	0.0
CD45RO CD4 lymphocyte act	22.5	Coronary artery SMC TNFalpha + IL-1beta	0.0
CD8 lymphocyte act	31.6	Astrocytes rest	0.0
Secondary CD8 lymphocyte rest	5.4	Astrocytes TNFalpha + IL-1beta	0.0
Secondary CD8 lymphocyte act	10.6	KU-812 (Basophil) rest	0.0
CD4 lymphocyte none	10.6	KU-812 (Basophil) PMA/ionomycin	0.0
2ry Th1/Th2/Tr1_anti-CD95 CH11	10.5	CCD1106 (Keratinocytes) none	0.0
LAK cells rest	4.7	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	0.0
LAK cells IL-2	19.1	Liver cirrhosis	0.0
LAK cells IL-2+IL-12	56.3	NCI-H292 none	0.0
LAK cells IL-2+IFN gamma	28.3	NCI-H292 IL-4	0.0
LAK cells IL-2+ IL-18	33.4	NCI-H292 IL-9	0.0
LAK cells PMA/ionomycin	0.0	NCI-H292 IL-13	0.0
NK Cells IL-2 rest	40.9	NCI-H292 IFN gamma	0.0
Two Way MLR 3 day	13.9	HPAEC none	0.0
Two Way MLR 5 day	3.4	HPAEC TNF alpha + IL-1 beta	0.0
Two Way MLR 7 day	25.7	Lung fibroblast none	0.0
PBMC rest	4.9	Lung fibroblast TNF alpha + IL-1 beta	0.0

PBMC PWM	21.3	Lung fibroblast IL-4	0.0
PBMC PHA-L	17.6	Lung fibroblast IL-9	0.0
Ramos (B cell) none	4.7	Lung fibroblast IL-13	0.0
Ramos (B cell) ionomycin	10.6	Lung fibroblast IFN gamma	0.0
B lymphocytes PWM	5.4	Dermal fibroblast CCD1070 rest	3.9
B lymphocytes CD40L and IL-4	6.7	Dermal fibroblast CCD1070 TNF alpha	53.2
EOL-1 dbcAMP	31.4	Dermal fibroblast CCD1070 IL-1 beta	0.0
EOL-1 dbcAMP PMA/ionomycin	3.5	Dermal fibroblast IFN gamma	0.0
Dendritic cells none	0.0	Dermal fibroblast IL-4	0.0
Dendritic cells LPS	0.0	Dermal Fibroblasts rest	0.0
Dendritic cells anti-CD40	0.0	Neutrophils TNFa+LPS	0.0
Monocytes rest	0.0	Neutrophils rest	0.0
Monocytes LPS	0.7	Colon	0.0
Macrophages rest	0.0	Lung	5.3
Macrophages LPS	0.0	Thymus	19.6
HUVEC none	11.3	Kidney	3.2
HUVEC starved	0.0		

Table AJD. Panel 5 Islet

Tissue Name	Rel. Exp.(%) Ag4880, Run 296908323	Tissue Name	Rel. Exp.(%) Ag4880, Run 296908323
97457_Patient-02go_adipose	0.0	94709_Donor 2 AM - A_adipose	0.0
97476_Patient-07sk_skeletal muscle	0.0	94710_Donor 2 AM - B_adipose	0.0
97477_Patient-07ut_uterus	0.0	94711_Donor 2 AM - C_adipose	0.0
97478_Patient-07pl_placenta	0.0	94712_Donor 2 AD - A_adipose	0.0
99167_Bayer Patient 1	0.0	94713_Donor 2 AD - B_adipose	0.0
97482_Patient-08ut_uterus	0.0	94714_Donor 2 AD - C_adipose	0.0
97483_Patient-08pl_placenta	0.0	94742_Donor 3 U - A Mesenchymal Stem Cells	0.0
97486_Patient-	0.0	94743_Donor 3 U -	0.0

09sk_skeletal muscle		B_Mesenchymal Stem Cells	
97487_Patient-09ut_uterus	0.0	94730_Donor 3 AM - A_adipose	0.0
97488_Patient-09pl_placenta	0.0	94731_Donor 3 AM - B_adipose	0.0
97492_Patient-10ut_uterus	0.0	94732_Donor 3 AM - C_adipose	0.0
97493_Patient-10pl_placenta	0.0	94733_Donor 3 AD - A_adipose	0.0
97495_Patient-11go_adipose	0.0	94734_Donor 3 AD - B_adipose	0.0
97496_Patient-11sk_skeletal muscle	0.0	94735_Donor 3 AD - C_adipose	0.0
97497_Patient-11ut_uterus	0.0	77138_Liver_HepG2untreated	0.0
97498_Patient-11pl_placenta	0.0	73556_Heart_Cardiac stromal cells (primary)	0.0
97500_Patient-12go_adipose	0.0	81735_Small Intestine	100.0
97501_Patient-12sk_skeletal muscle	0.0	72409_Kidney_Proximal Convoluted Tubule	0.0
97502_Patient-12ut_uterus	0.0	82685_Small intestine Duodenum	0.0
97503_Patient-12pl_placenta	0.0	90650_Adrenal_Adrenocortical adenoma	0.0
94721_Donor 2 U - A_Mesenchymal Stem Cells	0.0	72410_Kidney_HRCE	0.0
94722_Donor 2 U - B_Mesenchymal Stem Cells	0.0	72411_Kidney_HRE	0.0
94723_Donor 2 U - C_Mesenchymal Stem Cells	0.0	73139_Uterus_Uterine smooth muscle cells	0.0

**General\_screening\_panel\_v1.5 Summary:** Ag4880 Expression of this gene is highest in salivary gland (CT=20.3). Thus expression of this gene could be used to differentiate between this sample and other samples on this panel and as a marker of this tissue.

**Panel 4.1D Summary:** Ag4880 Highest expression of this gene is seen a sample derived from chronically activated Th1 cells (CT=32.2). Low but significant expression is seen in primary activated Th1 and Th2 cells, LAK cells, NK cells, eosinophils, TNF-a activated

dermal fibroblasts and thymus. This expression profile suggests that this gene product may be involved in autoimmune disease.

**Panel 5 Islet Summary:** Ag4880 Expression of this gene is limited to the small intestine (CT=23.7). Thus expression of this gene could be used to differentiate between this sample and other samples on this panel and as a marker of this tissue.

#### AK. CG160131-01: GLYCEROL KINASE.

Expression of gene CG160131-01 was assessed using the primer-probe set Ag5581, described in Table AKA. Results of the RTQ-PCR runs are shown in Tables AKB, AKC, AKD, AKE, AKF, AKG and AKH.

10 Table AKA. Probe Name Ag5581

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-accactgtagtctgggacaaga-3'	22	292	582
Probe	TET-5'-tctacaatgctgtggtgctccagtt-3'-TAMRA	26	329	583
Reverse	5'-acggcaactggaactgaag-3'	19	365	584

Table AKB. AI\_comprehensive panel\_v1.0

Tissue Name	Rel. Exp.(%) Ag5581, Run 244333633	Rel. Exp.(%) Ag5581, Run 244899563	Tissue Name	Rel. Exp.(%) Ag5581, Run 244333633	Rel. Exp.(%) Ag5581, Run 244899563
110967 COPD-F	0.0	0.0	112427 Match Control Psoriasis-F	0.0	6.7
110980 COPD-F	0.0	0.0	112418 Psoriasis-M	0.0	0.0
110968 COPD-M	3.9	0.0	112723 Match Control Psoriasis-M	0.0	0.0
110977 COPD-M	0.0	9.0	112419 Psoriasis-M	0.0	0.0
110989 Emphysema-F	0.0	7.4	112424 Match Control Psoriasis-M	3.4	4.1
110992 Emphysema-F	0.0	0.0	112420 Psoriasis-M	12.0	8.2
110993	4.2	0.0	112425 Match	0.0	0.0

Emphysema-F			Control Psoriasis-M		
110994 Emphysema-F	0.0	0.0	104689 (MF) OA Bone- Backus	13.9	13.5
110995 Emphysema-F	14.0	3.6	104690 (MF) Adj "Normal" Bone-Backus	0.0	15.8
110996 Emphysema-F	0.0	0.0	104691 (MF) OA Synovium- Backus	4.5	0.0
110997 Asthma-M	3.9	13.3	104692 (BA) OA Cartilage- Backus	0.0	0.0
111001 Asthma-F	0.0	0.0	104694 (BA) OA Bone- Backus	18.7	21.0
111002 Asthma-F	0.0	6.1	104695 (BA) Adj "Normal" Bone-Backus	0.0	8.4
111003 Atopic Asthma-F	4.3	0.0	104696 (BA) OA Synovium- Backus	23.7	15.5
111004 Atopic Asthma-F	0.0	0.0	104700 (SS) OA Bone- Backus	3.7	8.6
111005 Atopic Asthma-F	0.0	8.0	104701 (SS) Adj "Normal" Bone-Backus	5.6	27.5
111006 Atopic Asthma-F	0.0	0.0	104702 (SS) OA Synovium- Backus	7.3	0.0
111417 Allergy-M	0.0	0.0	117093 OA Cartilage Rep7	0.0	0.0
112347 Allergy-M	0.0	0.0	112672 OA Bone5	7.6	3.8
112349 Normal Lung-F	0.0	0.0	112673 OA Synovium5	7.6	7.7
112357 Normal Lung-F	0.0	0.0	112674 OA Synovial Fluid cells5	2.3	9.7
112354 Normal	0.0	0.0	117100 OA	0.0	0.0

Lung-M			Cartilage Rep14		
112374 Crohns-F	14.9	16.2	112756 OA Bone9	7.7	0.0
112389 Match Control Crohns-F	0.0	0.0	112757 OA Synovium9	10.6	9.7
112375 Crohns-F	0.0	4.5	112758 OA Synovial Fluid Cells9	0.0	0.0
112732 Match Control Crohns-F	0.0	6.2	117125 RA Cartilage Rep2	0.0	0.0
112725 Crohns-M	0.0	0.0	113492 Bone2 RA	66.0	40.9
112387 Match Control Crohns-M	0.0	7.6	113493 Synovium2 RA	7.5	7.5
112378 Crohns-M	0.0	0.0	113494 Syn Fluid Cells RA	23.3	46.0
112390 Match Control Crohns-M	5.5	7.1	113499 Cartilage4 RA	13.6	33.4
112726 Crohns-M	1.8	3.8	113500 Bone4 RA	68.8	37.1
112731 Match Control Crohns-M	1.3	7.7	113501 Synovium4 RA	29.9	54.3
112380 Ulcer Col-F	3.9	8.3	113502 Syn Fluid Cells4 RA	3.8	28.3
112734 Match Control Ulcer Col-F	100.0	100.0	113495 Cartilage3 RA	37.9	68.3
112384 Ulcer Col-F	3.7	0.0	113496 Bone3 RA	23.3	30.4
112737 Match Control Ulcer Col-F	0.0	0.0	113497 Synovium3 RA	27.7	0.0
112386 Ulcer Col-F	4.2	0.0	113498 Syn Fluid Cells3 RA	52.9	82.9
112738 Match Control Ulcer Col-F	15.5	66.0	117106 Normal Cartilage	0.0	0.0

			Rep20		
112381 Ulcer Col-M	0.0	0.0	113663 Bone3 Normal	8.1	0.0
112735 Match Control Ulcer Col-M	17.9	9.3	113664 Synovium3 Normal	0.0	0.0
112382 Ulcer Col-M	3.2	0.0	113665 Syn Fluid Cells3 Normal	0.0	0.0
112394 Match Control Ulcer Col-M	0.0	0.0	117107 Normal Cartilage Rep22	0.0	0.0
112383 Ulcer Col-M	1.3	0.0	113667 Bone4 Normal	4.6	0.0
112736 Match Control Ulcer Col-M	0.0	11.9	113668 Synovium4 Normal	0.0	8.9
112423 Psoriasis-F	10.6	22.1	113669 Syn Fluid Cells4 Normal	0.0	0.0

Table AKC. General\_screening\_panel\_v1.5.

Tissue Name	Rel. Exp.(%) Ag5581, Run 244896891	Tissue Name	Rel. Exp.(%) Ag5581, Run 244896891
Adipose	1.9	Renal ca. TK-10	22.2
Melanoma* Hs688(A).T	0.0	Bladder	22.1
Melanoma* Hs688(B).T	0.0	Gastric ca. (liver met.) NCI-N87	0.0
Melanoma* M14	0.0	Gastric ca. KATO III	1.6
Melanoma* LOXIMVI	0.0	Colon ca. SW-948	1.3
Melanoma* SK-MEL-5	2.0	Colon ca. SW480	1.2
Squamous cell carcinoma SCC-4	0.0	Colon ca.* (SW480 met) SW620	0.0
Testis Pool	0.0	Colon ca. HT29	0.0
Prostate ca.* (bone met) PC-3	0.5	Colon ca. HCT-116	0.0
Prostate Pool	0.6	Colon ca. CaCo-2	6.0
Placenta	0.7	Colon cancer tissue	27.2
Uterus Pool	0.0	Colon ca. SW1116	0.0

Ovarian ca. OVCAR-3	0.4	Colon ca. Colo-205	0.7
Ovarian ca. SK- OV-3	0.0	Colon ca. SW-48	0.6
Ovarian ca. OVCAR-4	0.0	Colon Pool	0.0
Ovarian ca. OVCAR-5	0.0	Small Intestine Pool	1.5
Ovarian ca. IGROV-1	2.0	Stomach Pool	0.5
Ovarian ca. OVCAR-8	3.4	Bone Marrow Pool	0.6
Ovary	0.0	Fetal Heart	0.8
Breast ca. MCF-7	0.5	Heart Pool	0.0
Breast ca. MDA- MB-231	0.7	Lymph Node Pool	0.6
Breast ca. BT 549	0.2	Fetal Skeletal Muscle	0.0
Breast ca. T47D	0.0	Skeletal Muscle Pool	15.9
Breast ca. MDA-N	2.4	Spleen Pool	0.6
Breast Pool	0.0	Thymus Pool	0.6
Trachea	3.3	CNS cancer (glio/astro) U87-MG	2.6
Lung	0.0	CNS cancer (glio/astro) U-118-MG	4.0
Fetal Lung	5.9	CNS cancer (neuro;met) SK-N-AS	0.0
Lung ca. NCI-N417	0.0	CNS cancer (astro) SF- 539	0.0
Lung ca. LX-1	0.0	CNS cancer (astro) SNB-75	2.4
Lung ca. NCI-H146	1.2	CNS cancer (glio) SNB-19	4.6
Lung ca. SHP-77	0.0	CNS cancer (glio) SF- 295	1.0
Lung ca. A549	0.0	Brain (Amygdala) Pool	1.7
Lung ca. NCI-H526	0.0	Brain (cerebellum)	3.7
Lung ca. NCI-H23	0.6	Brain (fetal)	0.0
Lung ca. NCI-H460	0.0	Brain (Hippocampus) Pool	7.6
Lung ca. HOP-62	0.0	Cerebral Cortex Pool	1.3
Lung ca. NCI-H522	0.0	Brain (Substantia nigra) Pool	3.9
Liver	8.1	Brain (Thalamus) Pool	1.5
Fetal Liver	100.0	Brain (whole)	4.2

Liver ca. HepG2	42.6	Spinal Cord Pool	15.1
Kidney Pool	1.1	Adrenal Gland	1.0
Fetal Kidney	2.2	Pituitary gland Pool	0.0
Renal ca. 786-0	0.5	Salivary Gland	0.7
Renal ca. A498	0.0	Thyroid (female)	1.0
Renal ca. ACHN	1.6	Pancreatic ca. CAPAN2	0.0
Renal ca. UO-31	0.0	Pancreas Pool	1.5

Table AKD. General\_screening\_panel\_v1.6

Tissue Name	Rel. Exp.(%) Ag5581, Run 278988931	Tissue Name	Rel. Exp.(%) Ag5581, Run 278988931
Adipose	6.1	Renal ca. TK-10	14.8
Melanoma* Hs688(A).T	0.0	Bladder	27.9
Melanoma* Hs688(B).T	0.0	Gastric ca. (liver met.) NCI-N87	1.7
Melanoma* M14	3.8	Gastric ca. KATO III	1.2
Melanoma* LOXIMVI	0.9	Colon ca. SW-948	0.0
Melanoma* SK- MEL-5	3.8	Colon ca. SW480	1.2
Squamous cell carcinoma SCC-4	0.0	Colon ca.* (SW480 met) SW620	0.9
Testis Pool	0.0	Colon ca. HT29	0.0
Prostate ca.* (bone met) PC-3	2.7	Colon ca. HCT-116	0.0
Prostate Pool	2.5	Colon ca. CaCo-2	5.7
Placenta	0.0	Colon cancer tissue	20.0
Uterus Pool	0.0	Colon ca. SW1116	0.0
Ovarian ca. OVCAR-3	0.7	Colon ca. Colo-205	0.0
Ovarian ca. SK- OV-3	0.7	Colon ca. SW-48	0.0
Ovarian ca. OVCAR-4	0.0	Colon Pool	0.0
Ovarian ca. OVCAR-5	0.0	Small Intestine Pool	1.6
Ovarian ca. IGROV-1	1.6	Stomach Pool	3.7
Ovarian ca. OVCAR-8	3.4	Bone Marrow Pool	0.0

Ovary	0.9	Fetal Heart	2.7
Breast ca. MCF-7	0.0	Heart Pool	1.4
Breast ca. MDA-MB-231	0.0	Lymph Node Pool	0.3
Breast ca. BT 549	0.0	Fetal Skeletal Muscle	1.0
Breast ca. T47D	0.8	Skeletal Muscle Pool	2.8
Breast ca. MDA-N	0.8	Spleen Pool	3.8
Breast Pool	0.6	Thymus Pool	1.6
Trachea	5.1	CNS cancer (glio/astro) U87-MG	1.9
Lung	0.0	CNS cancer (glio/astro) U-118-MG	3.2
Fetal Lung	7.0	CNS cancer (neuro;met) SK-N-AS	0.0
Lung ca. NCI-N417	0.0	CNS cancer (astro) SF-539	0.0
Lung ca. LX-1	0.0	CNS cancer (astro) SNB-75	0.8
Lung ca. NCI-H146	0.0	CNS cancer (glio) SNB-19	2.6
Lung ca. SHP-77	0.9	CNS cancer (glio) SF-295	2.8
Lung ca. A549	1.0	Brain (Amygdala) Pool	5.2
Lung ca. NCI-H526	0.0	Brain (cerebellum)	0.0
Lung ca. NCI-H23	0.0	Brain (fetal)	1.9
Lung ca. NCI-H460	0.8	Brain (Hippocampus) Pool	14.1
Lung ca. HOP-62	0.7	Cerebral Cortex Pool	6.5
Lung ca. NCI-H522	0.0	Brain (Substantia nigra) Pool	6.6
Liver	3.9	Brain (Thalamus) Pool	12.0
Fetal Liver	100.0	Brain (whole)	5.4
Liver ca. HepG2	29.9	Spinal Cord Pool	12.1
Kidney Pool	0.5	Adrenal Gland	4.2
Fetal Kidney	5.9	Pituitary gland Pool	0.8
Renal ca. 786-0	0.0	Salivary Gland	0.8
Renal ca. A498	0.0	Thyroid (female)	0.9
Renal ca. ACHN	0.0	Pancreatic ca. CAPAN2	0.0
Renal ca. UO-31	2.9	Pancreas Pool	0.0

Table AKE. Panel 4.1D.

Tissue Name	Rel. Exp.(%) Ag5581, Run 244337065	Tissue Name	Rel. Exp.(%) Ag5581, Run 244337065
Secondary Th1 act	0.1	HUVEC IL-1beta	0.0
Secondary Th2 act	0.2	HUVEC IFN gamma	0.0
Secondary Tr1 act	0.0	HUVEC TNF alpha + IFN gamma	0.0
Secondary Th1 rest	0.0	HUVEC TNF alpha + IL4	0.0
Secondary Th2 rest	0.0	HUVEC IL-11	0.0
Secondary Tr1 rest	0.0	Lung Microvascular EC none	0.0
Primary Th1 act	0.0	Lung Microvascular EC TNFalpha + IL-1beta	0.1
Primary Th2 act	0.2	Microvascular Dermal EC none	0.0
Primary Tr1 act	0.0	Microvascular Dermal EC TNFalpha + IL-1beta	0.0
Primary Th1 rest	0.0	Bronchial epithelium TNFalpha + IL1beta	0.0
Primary Th2 rest	0.0	Small airway epithelium none	0.0
Primary Tr1 rest	0.0	Small airway epithelium TNFalpha + IL-1beta	0.0
CD45RA CD4 lymphocyte act	0.1	Coronary artery SMC rest	0.0
CD45RO CD4 lymphocyte act	0.3	Coronary artery SMC TNFalpha + IL-1beta	0.0
CD8 lymphocyte act	0.0	Astrocytes rest	0.0
Secondary CD8 lymphocyte rest	0.1	Astrocytes TNFalpha + IL-1beta	0.0
Secondary CD8 lymphocyte act	0.0	KU-812 (Basophil) rest	0.0
CD4 lymphocyte none	0.0	KU-812 (Basophil) PMA/ionomycin	0.3
2ry Th1/Th2/Tr1_anti- CD95 CH11	0.0	CCD1106 (Keratinocytes) none	0.0
LAK cells rest	2.3	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	0.0
LAK cells IL-2	0.0	Liver cirrhosis	0.3
LAK cells IL-2+IL-12	0.0	NCI-H292 none	0.0
LAK cells IL-2+IFN gamma	0.0	NCI-H292 IL-4	0.2
LAK cells IL-2+ IL-18	0.1	NCI-H292 IL-9	0.1

LAK cells PMA/ionomycin	21.2	NCI-H292 IL-13	0.1
NK Cells IL-2 rest	0.0	NCI-H292 IFN gamma	0.0
Two Way MLR 3 day	1.9	HPAEC none	0.0
Two Way MLR 5 day	0.1	HPAEC TNF alpha + IL-1 beta	0.3
Two Way MLR 7 day	0.1	Lung fibroblast none	0.1
PBMC rest	0.0	Lung fibroblast TNF alpha + IL-1 beta	0.5
PBMC PWM	0.1	Lung fibroblast IL-4	0.1
PBMC PHA-L	0.5	Lung fibroblast IL-9	0.1
Ramos (B cell) none	0.0	Lung fibroblast IL-13	0.0
Ramos (B cell) ionomycin	0.0	Lung fibroblast IFN gamma	0.9
B lymphocytes PWM	0.1	Dermal fibroblast CCD1070 rest	0.2
B lymphocytes CD40L and IL-4	0.1	Dermal fibroblast CCD1070 TNF alpha	0.1
EOL-1 dbcAMP	0.0	Dermal fibroblast CCD1070 IL-1 beta	0.1
EOL-1 dbcAMP PMA/ionomycin	0.0	Dermal fibroblast IFN gamma	0.1
Dendritic cells none	2.3	Dermal fibroblast IL-4	0.1
Dendritic cells LPS	1.5	Dermal Fibroblasts rest	0.1
Dendritic cells anti-CD40	0.3	Neutrophils TNFa+LPS	21.2
Monocytes rest	0.0	Neutrophils rest	2.1
Monocytes LPS	100.0	Colon	0.1
Macrophages rest	1.0	Lung	0.0
Macrophages LPS	1.5	Thymus	0.0
HUVEC none	0.0	Kidney	1.5
HUVEC starved	0.0		

Table AKF. Panel 5 Islet

Tissue Name	Rel. Exp.(%) Ag5581, Run 244908254	Rel. Exp.(%) Ag5581, Run 279370998	Tissue Name	Rel. Exp.(%) Ag5581, Run 244908254	Rel. Exp.(%) Ag5581, Run 279370998
97457_Patient-02go_adipose	0.0	3.1	94709_Donor 2 AM - A_adipose	0.0	0.0
97476_Patient-07sk_skeletal	4.0	0.0	94710_Donor 2 AM - B_adipose	0.0	2.1

muscle					
97477_Patient-07ut_uterus	0.0	0.0	94711_Donor 2 AM - C_adipose	0.0	0.0
97478_Patient-07pl_placenta	5.1	3.3	94712_Donor 2 AD - A_adipose	0.0	0.0
99167_Bayer Patient 1	3.3	0.0	94713_Donor 2 AD - B_adipose	0.0	0.0
97482_Patient-08ut_uterus	4.3	0.0	94714_Donor 2 AD - C_adipose	0.0	0.0
97483_Patient-08pl_placenta	0.0	7.0	94742_Donor 3 U - A_Mesenchymal Stem Cells	0.0	0.0
97486_Patient-09sk_skeletal muscle	0.0	3.1	94743_Donor 3 U - B_Mesenchymal Stem Cells	0.0	0.0
97487_Patient-09ut_uterus	0.0	0.0	94730_Donor 3 AM - A_adipose	0.0	0.0
97488_Patient-09pl_placenta	0.0	0.0	94731_Donor 3 AM - B_adipose	0.0	0.0
97492_Patient-10ut_uterus	0.0	0.0	94732_Donor 3 AM - C_adipose	0.0	0.0
97493_Patient-10pl_placenta	0.0	3.7	94733_Donor 3 AD - A_adipose	0.0	0.0
97495_Patient-11go_adipose	0.0	2.3	94734_Donor 3 AD - B_adipose	0.0	0.0
97496_Patient-11sk_skeletal muscle	18.3	1.7	94735_Donor 3 AD - C_adipose	0.0	2.9
97497_Patient-11ut_uterus	0.0	2.1	77138_Liver_HepG2untreated	100.0	100.0
97498_Patient-11pl_placenta	0.0	0.0	73556_Heart_Cardiac stromal cells (primary)	0.0	0.0
97500_Patient-12go_adipose	0.0	0.0	81735_Small Intestine	35.4	29.7
97501_Patient-12sk_skeletal muscle	6.3	0.0	72409_Kidney_Proximal Convoluted Tubule	0.0	0.0
97502_Patient-12ut_uterus	0.0	0.0	82685_Small intestine_Duodenum	12.8	44.4
97503_Patient-12pl_placenta	0.0	6.6	90650_Adrenal_Adrenocortical adenoma	0.0	0.0
94721_Donor 2 U - A_Mesenchymal Stem Cells	0.0	0.0	72410_Kidney_HRCE	5.5	3.7
94722_Donor 2	0.0	0.0	72411_Kidney_HRE	0.0	0.0

U - B_Mesenchymal Stem Cells					
94723_Donor 2 U - C_Mesenchymal Stem Cells	0.0	7.2	73139_Uterus_Uterine smooth muscle cells	0.0	0.0

Table AKG. Panel 5D

Tissue Name	Rel. Exp.(%) Ag5581, Run 244988601	Tissue Name	Rel. Exp.(%) Ag5581, Run 244988601
97457_Patient- 02go_adipose	7.0	94709_Donor 2 AM - A_adipose	0.0
97476_Patient- 07sk_skeletal muscle	0.0	94710_Donor 2 AM - B_adipose	0.0
97477_Patient- 07ut_uterus	0.0	94711_Donor 2 AM - C_adipose	0.0
97478_Patient- 07pl_placenta	3.4	94712_Donor 2 AD - A_adipose	0.0
97481_Patient- 08sk_skeletal muscle	4.2	94713_Donor 2 AD - B_adipose	0.0
97482_Patient- 08ut_uterus	3.0	94714_Donor 2 AD - C_adipose	0.0
97483_Patient- 08pl_placenta	0.0	94742_Donor 3 U - A_Mesenchymal Stem Cells	0.0
97486_Patient- 09sk_skeletal muscle	0.0	94743_Donor 3 U - B_Mesenchymal Stem Cells	0.0
97487_Patient- 09ut_uterus	0.0	94730_Donor 3 AM - A_adipose	0.0
97488_Patient- 09pl_placenta	0.0	94731_Donor 3 AM - B_adipose	0.0
97492_Patient- 10ut_uterus	9.0	94732_Donor 3 AM - C_adipose	0.0
97493_Patient- 10pl_placenta	8.8	94733_Donor 3 AD - A_adipose	0.0
97495_Patient- 11go_adipose	4.9	94734_Donor 3 AD - B_adipose	0.0
97496_Patient- 11sk_skeletal muscle	0.0	94735_Donor 3 AD - C_adipose	0.0
97497_Patient- 11ut_uterus	0.0	77138_Liver_HepG2untreated	100.0
97498_Patient- 11pl_placenta	4.4	73556_Heart_Cardiac stromal cells (primary)	0.0

97500_Patient-12go_adipose	0.0	81735_Small Intestine	25.0
97501_Patient-12sk_skeletal muscle	4.9	72409_Kidney_Proximal Convoluted Tubule	0.0
97502_Patient-12ut_uterus	4.0	82685_Small intestine_Duodenum	40.3
97503_Patient-12pl_placenta	9.3	90650_Adrenal_Adrenocortical adenoma	0.0
94721_Donor 2 U - A_Mesenchymal Stem Cells	0.0	72410_Kidney_HRCE	3.3
94722_Donor 2 U - B_Mesenchymal Stem Cells	0.0	72411_Kidney_HRE	0.0
94723_Donor 2 U - C_Mesenchymal Stem Cells	0.0	73139_Uterus_Uterine smooth muscle cells	0.0

Table AKH. general oncology screening panel\_v\_2.4

Tissue Name	Rel. Exp.(%) Ag5581, Run 260268963	Tissue Name	Rel. Exp.(%) Ag5581, Run 260268963
Colon cancer 1	17.7	Bladder cancer NAT 2	0.0
Colon cancer NAT 1	0.0	Bladder cancer NAT 3	0.0
Colon cancer 2	15.4	Bladder cancer NAT 4	0.0
Colon cancer NAT 2	8.2	Prostate adenocarcinoma 1	2.7
Colon cancer 3	13.2	Prostate adenocarcinoma 2	0.0
Colon cancer NAT 3	6.1	Prostate adenocarcinoma 3	0.0
Colon malignant cancer 4	44.1	Prostate adenocarcinoma 4	2.2
Colon normal adjacent tissue 4	2.2	Prostate cancer NAT 5	0.0
Lung cancer 1	25.0	Prostate adenocarcinoma 6	0.0
Lung NAT 1	3.3	Prostate adenocarcinoma 7	3.3
Lung cancer 2	32.8	Prostate adenocarcinoma 8	0.0
Lung NAT 2	6.7	Prostate	0.0

		adenocarcinoma 9	
Squamous cell carcinoma 3	25.0	Prostate cancer NAT 10	0.0
Lung NAT 3	3.2	Kidney cancer 1	32.5
metastatic melanoma 1	1.5	Kidney NAT 1	2.9
Melanoma 2	1.2	Kidney cancer 2	12.4
Melanoma 3	0.0	Kidney NAT 2	10.7
metastatic melanoma 4	2.6	Kidney cancer 3	15.9
metastatic melanoma 5	14.2	Kidney NAT 3	16.4
Bladder cancer 1	6.2	Kidney cancer 4	12.9
Bladder cancer NAT 1	0.0	Kidney NAT 4	100.0
Bladder cancer 2	0.0		

**AI\_comprehensive\_panel\_v1.0 Summary:** Ag5581 Two experiments with the same probe and primer set show detectable expression of this gene limited to a sample of normal tissue adjacent to ulcerative colitis (CTs=33.5-34.5) and a sample derived from RA synovial fluid.

- 5 **General\_screening\_panel\_v1.5 Summary:** Ag5581 Highest expression is seen in fetal liver (CT=30.6). In addition, this gene is expressed at much higher levels in fetal liver tissue when compared to expression in the adult counterpart (CT=34). Thus, expression of this gene may be used to differentiate between the fetal and adult source of this tissue.

- 10 **General\_screening\_panel\_v1.6 Summary:** Ag5581 Highest expression is seen in fetal liver (CT=30.3). Overall, expression is in agreement with Panel 1.5. Please see that panel for further discussion of expression and utility of this gene.

- 15 **Panel 4.1D Summary:** Ag5581 Highest expression is seen in LPS treated monocytes (CT=27.4). Moderate levels of expression are seen in TFN-a/LPS treated neutrophils and PMA/ionomycin treated LAKs. Low but significant levels of expression are seen in macrophages. Upon activation with pathogens such as LPS, monocytes contribute to the innate and specific immunity by migrating to the site of tissue injury and releasing inflammatory cytokines. This release contributes to the inflammation process. Therefore expression of this gene could be used as a marker of activated monocytes. Furthermore,

modulation of the expression of the protein encoded by this transcript may prevent the recruitment of monocytes and the initiation of the inflammatory process, and reduce the symptoms of patients suffering from autoimmune and inflammatory diseases such as asthma, allergies, inflammatory bowel disease, lupus erythematosus, or rheumatoid arthritis.

**Panel 5 Islet Summary:** Ag5581 Two experiments with the same probe and primer set show detectable expression of this gene limited to a liver cancer cell line sample (CTs=33.5-34.5). This expression is in agreement with expression seen in Panels 1.5 and 1.6.

**Panel 5D Summary:** Ag5581 Expression of this gene limited to a liver cancer cell line sample (CT=34). This expression is in agreement with expression seen in Panels 1.5 and 1.6.

**General oncology screening panel\_v\_2.4 Summary:** Ag5581 Highest expression is seen in a kidney sample (CT=32). In addition, this gene is more highly expressed in lung and colon cancer than in the corresponding normal adjacent tissue. Thus, expression of this gene could be used as a marker of these cancers. Furthermore, therapeutic modulation of the expression or function of this gene product may be useful in the treatment of lung and colon cancer.

#### AL. CG160131-04: FL\_1\_552 GLYCEROL KINASE.

Expression of gene CG160131-04 was assessed using the primer-probe set Ag7439, described in Table ALA. Results of the RTQ-PCR runs are shown in Tables ALB and ALC. Please note that CG160131-04 represents a full-length physical clone.

Table ALA. Probe Name Ag7439

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-agcacatttggtccaccaat-3'	20	774	585
Probe	TET-5'-caccagatattggcacaccttccaa-3'-TAMRA	26	815	586
Reverse	5'-atgaaaatctctcatagcgtgaa-3'	23	851	587

Table ALB. AI\_comprehensive panel\_v1.0

<b>Tissue Name</b>	<b>Rel. Exp.(%) Ag7439, Run 311756513</b>	<b>Tissue Name</b>	<b>Rel. Exp.(%) Ag7439, Run 311756513</b>
110967 COPD-F	19.1	112427 Match Control Psoriasis-F	100.0
110980 COPD-F	18.3	112418 Psoriasis-M	23.5
110968 COPD-M	16.5	112723 Match Control Psoriasis-M	21.2
110977 COPD-M	68.8	112419 Psoriasis-M	43.8
110989 Emphysema-F	54.3	112424 Match Control Psoriasis-M	23.2
110992 Emphysema-F	8.2	112420 Psoriasis-M	79.6
110993 Emphysema-F	35.8	112425 Match Control Psoriasis-M	82.9
110994 Emphysema-F	13.5	104689 (MF) OA Bone-Backus	20.0
110995 Emphysema-F	29.3	104690 (MF) Adj "Normal" Bone- Backus	24.0
110996 Emphysema-F	2.1	104691 (MF) OA Synovium-Backus	71.7
110997 Asthma-M	2.5	104692 (BA) OA Cartilage-Backus	0.0
111001 Asthma-F	32.8	104694 (BA) OA Bone-Backus	27.2
111002 Asthma-F	26.2	104695 (BA) Adj "Normal" Bone- Backus	24.3
111003 Atopic Asthma-F	30.6	104696 (BA) OA Synovium-Backus	57.4
111004 Atopic Asthma-F	18.6	104700 (SS) OA Bone-Backus	16.2
111005 Atopic Asthma-F	17.6	104701 (SS) Adj "Normal" Bone- Backus	18.2
111006 Atopic Asthma-F	4.1	104702 (SS) OA Synovium-Backus	39.8
111417 Allergy-M	12.2	117093 OA Cartilage Rep7	31.4
112347 Allergy-M	0.0	112672 OA Bone5	77.4
112349 Normal Lung-F	0.0	112673 OA Synovium5	35.8
112357 Normal Lung-F	52.1	112674 OA Synovial Fluid cells5	47.0

112354 Normal Lung-M	27.7	117100 OA Cartilage Rep14	8.4
112374 Crohns-F	27.2	112756 OA Bone9	69.3
112389 Match Control Crohns-F	8.6	112757 OA Synovium9	20.6
112375 Crohns-F	20.7	112758 OA Synovial Fluid Cells9	9.2
112732 Match Control Crohns-F	4.7	117125 RA Cartilage Rep2	13.4
112725 Crohns-M	5.4	113492 Bone2 RA	18.7
112387 Match Control Crohns-M	12.6	113493 Synovium2 RA	4.2
112378 Crohns-M	0.0	113494 Syn Fluid Cells RA	6.8
112390 Match Control Crohns-M	56.3	113499 Cartilage4 RA	7.7
112726 Crohns-M	21.2	113500 Bone4 RA	11.0
112731 Match Control Crohns-M	20.0	113501 Synovium4 RA	9.2
112380 Ulcer Col-F	31.9	113502 Syn Fluid Cells4 RA	4.5
112734 Match Control Ulcer Col-F	15.3	113495 Cartilage3 RA	6.3
112384 Ulcer Col-F	43.2	113496 Bone3 RA	7.2
112737 Match Control Ulcer Col-F	5.5	113497 Synovium3 RA	4.6
112386 Ulcer Col-F	15.2	113498 Syn Fluid Cells3 RA	10.3
112738 Match Control Ulcer Col-F	5.6	117106 Normal Cartilage Rep20	2.8
112381 Ulcer Col-M	0.1	113663 Bone3 Normal	0.0
112735 Match Control Ulcer Col-M	3.0	113664 Synovium3 Normal	0.0
112382 Ulcer Col-M	18.4	113665 Syn Fluid Cells3 Normal	0.0
112394 Match Control Ulcer Col-M	8.9	117107 Normal Cartilage Rep22	5.8
112383 Ulcer Col-M	24.7	113667 Bone4 Normal	32.5
112736 Match Control Ulcer Col-M	6.3	113668 Synovium4 Normal	21.8
112423 Psoriasis-F	21.8	113669 Syn Fluid Cells4 Normal	43.2

Table ALC. Panel 4.1D.

Tissue Name	Rel. Exp.(%) Ag7439, Run 305901963	Tissue Name	Rel. Exp.(%) Ag7439, Run 305901963
Secondary Th1 act	1.9	HUVEC IL-1beta	2.2
Secondary Th2 act	1.7	HUVEC IFN gamma	1.8
Secondary Tr1 act	1.0	HUVEC TNF alpha + IFN gamma	0.6
Secondary Th1 rest	0.0	HUVEC TNF alpha + IL4	1.7
Secondary Th2 rest	0.3	HUVEC IL-11	0.6
Secondary Tr1 rest	0.0	Lung Microvascular EC none	3.4
Primary Th1 act	0.1	Lung Microvascular EC TNFalpha + IL-1beta	3.2
Primary Th2 act	1.2	Microvascular Dermal EC none	0.1
Primary Tr1 act	2.0	Microvascular Dermal EC TNFalpha + IL-1beta	1.3
Primary Th1 rest	0.0	Bronchial epithelium TNFalpha + IL1beta	0.4
Primary Th2 rest	0.0	Small airway epithelium none	0.3
Primary Tr1 rest	0.0	Small airway epithelium TNFalpha + IL-1beta	0.0
CD45RA CD4 lymphocyte act	5.2	Coronary artery SMC rest	2.2
CD45RO CD4 lymphocyte act	1.1	Coronary artery SMC TNFalpha + IL-1beta	3.6
CD8 lymphocyte act	0.0	Astrocytes rest	0.0
Secondary CD8 lymphocyte rest	4.0	Astrocytes TNFalpha + IL-1beta	0.8
Secondary CD8 lymphocyte act	0.0	KU-812 (Basophil) rest	0.3
CD4 lymphocyte none	0.0	KU-812 (Basophil) PMA/ionomycin	0.1
2ry Th1/Th2/Tr1_anti- CD95 CH11	0.5	CCD1106 (Keratinocytes) none	0.0
LAK cells rest	0.3	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	0.0
LAK cells IL-2	0.2	Liver cirrhosis	0.0
LAK cells IL-2+IL-12	0.0	NCI-H292 none	0.0
LAK cells IL-2+IFN gamma	0.2	NCI-H292 IL-4	0.2
LAK cells IL-2+ IL-18	0.2	NCI-H292 IL-9	0.0

LAK cells PMA/ionomycin	14.6	NCI-H292 IL-13	1.0
NK Cells IL-2 rest	0.9	NCI-H292 IFN gamma	0.0
Two Way MLR 3 day	3.1	HPAEC none	0.8
Two Way MLR 5 day	0.4	HPAEC TNF alpha + IL-1 beta	8.8
Two Way MLR 7 day	0.1	Lung fibroblast none	24.5
PBMC rest	0.1	Lung fibroblast TNF alpha + IL-1 beta	43.2
PBMC PWM	0.9	Lung fibroblast IL-4	14.5
PBMC PHA-L	0.5	Lung fibroblast IL-9	21.8
Ramos (B cell) none	0.1	Lung fibroblast IL-13	12.9
Ramos (B cell) ionomycin	0.1	Lung fibroblast IFN gamma	100.0
B lymphocytes PWM	0.4	Dermal fibroblast CCD1070 rest	6.1
B lymphocytes CD40L and IL-4	0.2	Dermal fibroblast CCD1070 TNF alpha	11.6
EOL-1 dbcAMP	1.9	Dermal fibroblast CCD1070 IL-1 beta	11.0
EOL-1 dbcAMP PMA/ionomycin	0.1	Dermal fibroblast IFN gamma	7.2
Dendritic cells none	1.7	Dermal fibroblast IL-4	3.8
Dendritic cells LPS	1.2	Dermal Fibroblasts rest	2.6
Dendritic cells anti- CD40	0.3	Neutrophils TNFa+LPS	2.6
Monocytes rest	0.1	Neutrophils rest	1.6
Monocytes LPS	31.2	Colon	0.0
Macrophages rest	0.4	Lung	0.3
Macrophages LPS	0.8	Thymus	0.2
HUVEC none	0.6	Kidney	1.1
HUVEC starved	2.8		

**AI\_comprehensive panel\_v1.0 Summary:** Ag7439 Highest expression is seen in normal tissue adjacent to psoriasis (CT=29.8). In addition, moderate to low levels of expression are seen in many samples on this panel. Thus, this gene product may be involved in autoimmune disease.

- 5 **CNS\_neurodegeneration\_v1.0 Summary:** Ag7439 Results from one experiment with this gene are not included. The amp plot indicates that there were experimental difficulties with this run.

**Panel 4.1D Summary:** Ag7439 Highest expression is seen in a sample of IFN gama lung derived fibroblasts (CT=29). Low but significant levels of expression are also seen in clusters of samples derived from lung and dermal fibroblasts. Thus, this gene product may be involved in inflammatory processes of the lung and skin, including psoriasis, asthma, emphysema, and allergy.

**Panel 5 Islet Summary:** Ag7439 Expression of this gene is low/undetectable in all samples on this panel (CTs>35). (Data not shown.)

**AM. CG166282-01: CHK1-variant.**

Expression of gene CG166282-01 was assessed using the primer-probe set Ag5448, described in Table AMA. Results of the RTQ-PCR runs are shown in Tables AMB, AMC and AMD.

Table AMA. Probe Name Ag5448

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-tgtatgaatcagggatgatggat-3'	22	1256	588
Probe	TET-5'-tcttcaggaagtgtctcttgaactcca-3'-TAMRA	27	1278	589
Reverse	5'-ctggctgctcacaatatcaatc-3'	22	1318	590

Table AMB. General\_screening\_panel\_v1.5

Tissue Name	Rel. Exp.(%) Ag5448, Run 237375423	Rel. Exp.(%) Ag5448, Run 247291071	Tissue Name	Rel. Exp.(%) Ag5448, Run 237375423	Rel. Exp.(%) Ag5448, Run 247291071
Adipose	0.2	0.0	Renal ca. TK-10	8.2	5.0
Melanoma* Hs688(A).T	6.7	3.1	Bladder	3.2	0.0
Melanoma* Hs688(B).T	5.0	4.3	Gastric ca. (liver met.) NCI-N87	8.0	8.0
Melanoma* M14	25.5	18.8	Gastric ca. KATO III	100.0	100.0
Melanoma* LOXIMVI	28.7	22.5	Colon ca. SW- 948	7.7	7.9
Melanoma* SK-MEL-5	17.3	12.2	Colon ca. SW480	62.0	46.0
Squamous cell	5.6	5.6	Colon ca.* (SW480 met)	32.8	31.9

carcinoma SCC-4			SW620		
Testis Pool	0.5	2.2	Colon ca. HT29	18.6	5.1
Prostate ca.* (bone met) PC-3	11.7	9.3	Colon ca. HCT- 116	33.9	39.5
Prostate Pool	0.0	0.0	Colon ca. CaCo- 2	27.0	19.3
Placenta	0.0	1.4	Colon cancer tissue	5.5	4.4
Uterus Pool	0.3	0.0	Colon ca. SW1116	4.1	5.3
Ovarian ca. OVCAR-3	10.2	6.5	Colon ca. Colo- 205	8.8	8.4
Ovarian ca. SK-OV-3	32.3	35.8	Colon ca. SW-48	13.6	8.0
Ovarian ca. OVCAR-4	22.8	16.5	Colon Pool	0.4	0.0
Ovarian ca. OVCAR-5	12.6	5.5	Small Intestine Pool	1.0	0.0
Ovarian ca. IGROV-1	8.7	9.5	Stomach Pool	0.6	0.0
Ovarian ca. OVCAR-8	10.4	9.0	Bone Marrow Pool	0.0	0.0
Ovary	0.2	0.0	Fetal Heart	2.4	2.6
Breast ca. MCF-7	4.2	5.8	Heart Pool	0.6	0.0
Breast ca. MDA-MB- 231	56.3	45.7	Lymph Node Pool	0.0	0.0
Breast ca. BT 549	27.9	16.7	Fetal Skeletal Muscle	0.3	0.0
Breast ca. T47D	17.6	15.0	Skeletal Muscle Pool	0.0	0.0
Breast ca. MDA-N	12.9	14.8	Spleen Pool	0.2	1.9
Breast Pool	0.1	0.0	Thymus Pool	2.1	2.4
Trachea	0.8	0.0	CNS cancer (glio/astro) U87- MG	14.2	9.2
Lung	0.0	0.0	CNS cancer (glio/astro) U- 118-MG	44.1	47.3
Fetal Lung	2.2	0.0	CNS cancer (neuro;met) SK-	8.3	12.2

			N-AS		
Lung ca. NCI-N417	8.1	7.2	CNS cancer (astro) SF-539	6.4	8.5
Lung ca. LX- 1	20.0	8.8	CNS cancer (astro) SNB-75	29.9	30.6
Lung ca. NCI-H146	9.7	9.3	CNS cancer (glio) SNB-19	4.4	5.4
Lung ca. SHP-77	22.5	16.0	CNS cancer (glio) SF-295	8.7	4.2
Lung ca. A549	18.7	10.5	Brain (Amygdala) Pool	0.0	0.0
Lung ca. NCI-H526	30.1	26.4	Brain (cerebellum)	0.5	0.0
Lung ca. NCI-H23	16.5	12.3	Brain (fetal)	2.7	1.2
Lung ca. NCI-H460	4.5	1.2	Brain (Hippocampus) Pool	0.6	0.0
Lung ca. HOP-62	2.6	2.2	Cerebral Cortex Pool	0.0	0.0
Lung ca. NCI-H522	32.8	31.2	Brain (Substantia nigra) Pool	0.0	0.0
Liver	0.0	0.0	Brain (Thalamus) Pool	0.4	0.0
Fetal Liver	4.5	6.9	Brain (whole)	0.0	0.0
Liver ca. HepG2	5.8	4.1	Spinal Cord Pool	0.1	0.0
Kidney Pool	0.4	0.0	Adrenal Gland	0.0	0.0
Fetal Kidney	6.2	7.4	Pituitary gland Pool	0.2	0.0
Renal ca. 786-0	12.5	12.5	Salivary Gland	0.4	0.0
Renal ca. A498	2.0	3.5	Thyroid (female)	0.2	0.0
Renal ca. ACHN	6.6	3.2	Pancreatic ca. CAPAN2	40.1	48.0
Renal ca. UO-31	23.3	21.3	Pancreas Pool	1.4	0.0

Table AMC. Panel 4.1D

Tissue Name	Rel. Exp.(%) Ag5448, Run 237371903	Tissue Name	Rel. Exp.(%) Ag5448, Run 237371903
Secondary Th1. act	88.9	HUVEC IL-1beta	44.1

Secondary Th2 act	100.0	HUVEC IFN gamma	17.1
Secondary Tr1 act	16.6	HUVEC TNF alpha + IFN gamma	2.2
Secondary Th1 rest	0.0	HUVEC TNF alpha + IL4	1.6
Secondary Th2 rest	0.0	HUVEC IL-11	15.5
Secondary Tr1 rest	0.0	Lung Microvascular EC none	11.9
Primary Th1 act	0.0	Lung Microvascular EC TNFalpha + IL-1beta	7.8
Primary Th2 act	47.3	Microvascular Dermal EC none	3.8
Primary Tr1 act	50.7	Microvascular Dermal EC TNFalpha + IL-1beta	3.7
Primary Th1 rest	1.5	Bronchial epithelium TNFalpha + IL1beta	3.1
Primary Th2 rest	7.9	Small airway epithelium none	12.7
Primary Tr1 rest	0.0	Small airway epithelium TNFalpha + IL-1beta	19.6
CD45RA CD4 lymphocyte act	41.5	Coronary artery SMC rest	7.5
CD45RO CD4 lymphocyte act	77.9	Coronary artery SMC TNFalpha + IL-1beta	2.2
CD8 lymphocyte act	11.0	Astrocytes rest	1.4
Secondary CD8 lymphocyte rest	99.3	Astrocytes TNFalpha + IL-1beta	0.0
Secondary CD8 lymphocyte act	13.3	KU-812 (Basophil) rest	34.6
CD4 lymphocyte none	0.0	KU-812 (Basophil) PMA/ionomycin	45.1
2ry Th1/Th2/Tr1_anti-CD95 CH11	0.0	CCD1106 (Keratinocytes) none	24.7
LAK cells rest	1.6	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	10.5
LAK cells IL-2	15.4	Liver cirrhosis	0.0
LAK cells IL-2+IL-12	2.0	NCI-H292 none	13.7
LAK cells IL-2+IFN gamma	17.6	NCI-H292 IL-4	38.7
LAK cells IL-2+ IL-18	5.6	NCI-H292 IL-9	23.7
LAK cells PMA/ionomycin	13.3	NCI-H292 IL-13	41.2
NK Cells IL-2 rest	35.1	NCI-H292 IFN gamma	22.8
Two Way MLR 3 day	1.2	HPAEC none	4.4

Two Way MLR 5 day	6.3	HPAEC TNF alpha + IL-1 beta	11.2
Two Way MLR 7 day	2.2	Lung fibroblast none	5.2
PBMC rest	0.0	Lung fibroblast TNF alpha + IL-1 beta	7.1
PBMC PWM	14.6	Lung fibroblast IL-4	1.6
PBMC PHA-L	7.3	Lung fibroblast IL-9	0.0
Ramos (B cell) none	4.0	Lung fibroblast IL-13	0.0
Ramos (B cell) ionomycin	41.5	Lung fibroblast IFN gamma	3.5
B lymphocytes PWM	45.4	Dermal fibroblast CCD1070 rest	33.9
B lymphocytes CD40L and IL-4	27.0	Dermal fibroblast CCD1070 TNF alpha	59.0
EOL-1 dbcAMP	74.2	Dermal fibroblast CCD1070 IL-1 beta	23.8
EOL-1 dbcAMP PMA/ionomycin	2.3	Dermal fibroblast IFN gamma	22.4
Dendritic cells none	0.0	Dermal fibroblast IL-4	31.0
Dendritic cells LPS	0.0	Dermal Fibroblasts rest	16.7
Dendritic cells anti-CD40	0.0	Neutrophils TNFa+LPS	0.0
Monocytes rest	0.0	Neutrophils rest	0.0
Monocytes LPS	11.3	Colon	0.0
Macrophages rest	0.0	Lung	0.0
Macrophages LPS	0.0	Thymus	0.0
HUVEC none	26.4	Kidney	0.0
HUVEC starved	24.0		

Table AMD. general oncology screening panel\_v\_2.4

Tissue Name	Rel. Exp.(%) Ag5448, Run 260285334	Tissue Name	Rel. Exp.(%) Ag5448, Run 260285334
Colon cancer 1	15.9	Bladder NAT 2	0.0
Colon NAT 1	3.4	Bladder NAT 3	0.0
Colon cancer 2	26.8	Bladder NAT 4	0.0
Colon NAT 2	15.6	Prostate adenocarcinoma 1	0.0
Colon cancer 3	51.8	Prostate adenocarcinoma 2	0.0
Colon NAT 3	3.6	Prostate adenocarcinoma 3	0.0
Colon malignant cancer 4	100.0	Prostate adenocarcinoma 4	3.2

Colon NAT 4	4.0	Prostate NAT 5	0.0
Lung cancer 1	8.3	Prostate adenocarcinoma 6	0.0
Lung NAT 1	0.0	Prostate adenocarcinoma 7	0.0
Lung cancer 2	33.9	Prostate adenocarcinoma 8	0.0
Lung NAT 2	0.0	Prostate adenocarcinoma 9	0.0
Squamous cell carcinoma 3	15.5	Prostate NAT 10	0.0
Lung NAT 3	0.0	Kidney cancer 1	0.0
Metastatic melanoma 1	0.0	Kidney NAT 1	0.0
Melanoma 2	0.0	Kidney cancer 2	15.9
Melanoma 3	0.0	Kidney NAT 2	0.0
Metastatic melanoma 4	5.1	Kidney cancer 3	5.2
Metastatic melanoma 5	3.8	Kidney NAT 3	0.0
Bladder cancer 1	0.0	Kidney cancer 4	0.0
Bladder NAT 1	0.0	Kidney NAT 4	0.0
Bladder cancer 2	4.4		

**AI\_comprehensive\_panel\_v1.0 Summary:** Ag5448 The amp plot indicates that there were experimental difficulties with this run; therefore, no conclusions can be drawn from this data. (Data not shown).

**General\_screening\_panel\_v1.5 Summary:** Ag5448 Two experiments with same probe-primer sets are in excellent agreement, with highest expression of this gene detected in gastric cancer KATO III cell line (CTs=30-33). Moderate to low levels of expression of this gene is also seen in cluster of cancer cell lines derived from pancreatic, gastric, colon, lung, liver, renal, breast, ovarian, prostate, squamous cell carcinoma, melanoma and brain cancers. Thus, expression of this gene could be used as a marker to detect the presence of these cancers. Furthermore, therapeutic modulation of the expression or function of this gene may be effective in the treatment of pancreatic, gastric, colon, lung, liver, renal, breast, ovarian, prostate, squamous cell carcinoma, melanoma and brain cancers.

**Oncology\_cell\_line\_screening\_panel\_v3.2 Summary:** Ag5448 The amp plot indicates that there were experimental difficulties with this run; therefore, no conclusions can be drawn from this data. (Data not shown).

- Panel 4.1D Summary:** Ag5448 Highest expression of this gene is detected in activated secondary Th2 cells (CT=33). Low expression of this gene is detected in activated polarized T cells, resting IL-2 treated NK cells, activated Ramos B cells and B lymphocytes, eosinophils, activated HUVEC cells and NCI-H292 cells, basophils and TNF alpha stimulated dermal fibroblasts. Therefore, therapeutic modulation of this gene product may ameliorate symptoms/conditions associated with autoimmune and inflammatory disorders including psoriasis, allergy, asthma, inflammatory bowel disease, rheumatoid arthritis and osteoarthritis.

- General oncology screening panel\_v\_2.4 Summary:** Ag5448 Highest expression of this gene malignant colon cancer (CT=34.4). Higher expression of this gene is associated with the colon cancer as compared to adjacent control tissue. Therefore, expression of this gene may be used as diagnostic marker to detect colon cancer and also, therapeutic modulation of this gene or its protein product may be useful in the treatment of colon cancer.

**AN. CG170739-01: PENDRIN.**

Expression of gene CG170739-01 was assessed using the primer-probe set Ag6134, described in Table ANA.

20. Table ANA. Probe Name Ag6134

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-cgctgcaaggaccttttc-3'	18	1931	591
Probe	TET-5'-tgctcagaacaacagatcccaccatt-3'-TAMRA	26	1892	592
Reverse	5'-tgctggatacgagaaagtgttc-3'	22	1859	593

**AI\_comprehensive\_panel\_v1.0 Summary:** Ag6134 Expression of this gene is low/undetectable (CTs > 35) across all of the samples on this panel (data not shown). The amp plot indicates that there is a high probability of a probe failure.

**General\_screening\_panel\_v1.5 Summary:** Ag6134 Expression of this gene is low/undetectable (CTs > 35) across all of the samples on this panel (data not shown). The amp plot indicates that there is a high probability of a probe failure.

**Panel 4.1D Summary:** Ag6134 Expression of this gene is low/undetectable (CTs > 35) across all of the samples on this panel (data not shown). The amp plot indicates that there is a high probability of a probe failure.

**AO. CG51213-07: CG51213-(13-364).**

Expression of gene CG51213-07 was assessed using the primer-probe sets Ag1425, Ag813, Ag871 and Ag924, described in Tables AOA, AOB, AOC and AOD. Results of the RTQ-PCR runs are shown in Tables AOE, AOF, AOG, AOH, AOI, AOJ and AOK.

Table AOA. Probe Name Ag1425

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-ggacttcagagaagtgcagtgt-3'	22	549	594
Probe	TET-5'-ctgaatttgacagcatccctttccgt-3'-TAMRA	26	572	595
Reverse	5'-cgggtacgttttccacttgtaga-3'	22	605	596

Table AOB. Probe Name Ag813

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-tgtagaatttcccacggaag-3'	21	590	597
Probe	TET-5'-cactgcacttctctgaagtcctggga-3'-TAMRA	26	544	598
Reverse	5'-ctgcaacacggatgactgt-3'	19	516	599

Table AOC. Probe Name Ag871

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-tctagctgggaccacctttc-3'	20	1041	600
Probe	TET-5'-cagaccaggtccagagcctcgaag-3'-TAMRA	24	1076	601
Reverse	5'-acgatgagagatgcattaatcg-3'	22	1109	602

Table AOD. Probe Name Ag924

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5' -ggacttcagagaagtgcagtgt-3'	22	549	603
Probe	TET-5' -ctgaatttgacagcatccctttccgt-3' - TAMRA	26	572	604
Reverse	5' -cggtagcggtttccacttgtaga-3'	22	605	605

Table AOE. AI\_comprehensive panel\_v1.0

Tissue Name	Rel. Exp.(%) Ag813, Run 234222162	Rel. Exp.(%) Ag813, Run 246953625	Tissue Name	Rel. Exp.(%) Ag813, Run 234222162	Rel. Exp.(%) Ag813, Run 246953625
110967 COPD-F	5.4	8.8	112427 Match Control Psoriasis-F	0.0	30.6
110980 COPD-F	5.9	9.2	112418 Psoriasis-M	8.6	8.7
110968 COPD-M	12.9	11.9	112723 Match Control Psoriasis-M	11.0	8.8
110977 COPD-M	18.8	25.7	112419 Psoriasis-M	10.7	8.1
110989 Emphysema-F	19.3	26.4	112424 Match Control Psoriasis-M	7.4	4.1
110992 Emphysema-F	13.5	30.8	112420 Psoriasis-M	37.4	36.3
110993 Emphysema-F	10.5	13.2	112425 Match Control Psoriasis-M	11.7	6.2
110994 Emphysema-F	10.4	7.3	104689 (MF) OA Bone- Backus	100.0	100.0
110995 Emphysema-F	25.5	25.9	104690 (MF) Adj "Normal" Bone-Backus	62.0	65.5
110996 Emphysema-F	3.7	6.5	104691 (MF) OA Synovium- Backus	73.7	74.7
110997 Asthma-M	2.5	2.4	104692 (BA) OA Cartilage- Backus	15.8	15.0
111001 Asthma-F	16.3	21.0	104694 (BA) OA Bone- Backus	69.3	79.0

111002 Asthma-F	24.0	22.1	104695 (BA) Adj "Normal" Bone-Backus	68.3	44.1
111003 Atopic Asthma-F	14.9	35.4	104696 (BA) OA Synovium- Backus	29.5	27.9
111004 Atopic Asthma-F	31.6	47.0	104700 (SS) OA Bone- Backus	55.1	43.2
111005 Atopic Asthma-F	18.4	20.2	104701 (SS) Adj "Normal" Bone-Backus	72.2	95.3
111006 Atopic Asthma-F	2.6	5.6	104702 (SS) OA Synovium- Backus	36.3	37.9
111417 Allergy-M	13.4	8.5	117093 OA Cartilage Rep7	4.9	11.3
112347 Allergy-M	0.0	0.0	112672 OA Bone5	25.3	25.0
112349 Normal Lung-F	0.0	0.0	112673 OA Synovium5	8.4	12.6
112357 Normal Lung-F	15.5	16.4	112674 OA Synovial Fluid cells5	18.8	16.2
112354 Normal Lung-M	3.7	1.5	117100 OA Cartilage Rep14	8.0	10.5
112374 Crohns-F	16.6	21.6	112756 OA Bone9	3.6	11.2
112389 Match Control Crohns-F	10.3	6.3	112757 OA Synovium9	6.0	5.4
112375 Crohns-F	0.0	32.8	112758 OA Synovial Fluid Cells9	9.9	9.4
112732 Match Control Crohns-F	10.4	9.7	117125 RA Cartilage Rep2	5.3	9.3
112725 Crohns-M	2.2	0.8	113492 Bone2 RA	4.0	4.1
112387 Match Control Crohns-M	8.4	10.5	113493 Synovium2 RA	1.0	1.7
112378	0.0	0.0	113494 Syn	2.6	5.6

Crohns-M			Fluid Cells RA		
112390 Match Control Crohns-M	38.7	38.2	113499 Cartilage4 RA	4.7	5.2
112726 Crohns-M	27.4	22.8	113500 Bone4 RA	4.0	4.6
112731 Match Control Crohns-M	7.6	13.6	113501 Synovium4 RA	3.6	3.1
112380 Ulcer Col-F	15.9	20.4	113502 Syn Fluid Cells4 RA	2.3	1.9
112734 Match Control Ulcer Col-F	13.5	26.4	113495 Cartilage3 RA	3.3	5.4
112384 Ulcer Col-F	21.6	18.8	113496 Bone3 RA	4.6	6.4
112737 Match Control Ulcer Col-F	5.6	5.8	113497 Synovium3 RA	3.1	1.6
112386 Ulcer Col-F	0.7	1.1	113498 Syn Fluid Cells3 RA	6.9	6.0
112738 Match Control Ulcer Col-F	3.0	4.3	117106 Normal Cartilage Rep20	13.7	13.0
112381 Ulcer Col-M	0.0	0.1	113663 Bone3 Normal	0.0	0.0
112735 Match Control Ulcer Col-M	2.1	0.8	113664 Synovium3 Normal	0.0	0.0
112382 Ulcer Col-M	8.7	8.5	113665 Syn Fluid Cells3 Normal	0.0	0.1
112394 Match Control Ulcer Col-M	1.5	4.7	117107 Normal Cartilage Rep22	2.3	0.3
112383 Ulcer Col-M	45.7	54.7	113667 Bone4 Normal	8.6	4.7
112736 Match Control Ulcer Col-M	6.1	6.4	113668 Synovium4 Normal	3.0	6.4
112423	7.7	5.2	113669 Syn	12.7	11.0

Psoriasis-F			Fluid Cells4 Normal		
-------------	--	--	------------------------	--	--

Table AOF. CNS\_neurodegeneration\_v1.0

Tissue Name	Rel. Exp.(%) Ag813, Run 209990454	Tissue Name	Rel. Exp.(%) Ag813, Run 209990454
AD 1 Hippo	43.5	Control (Path) 3 Temporal Ctx	32.1
AD 2 Hippo	50.0	Control (Path) 4 Temporal Ctx	57.0
AD 3 Hippo	35.4	AD 1 Occipital Ctx	60.7
AD 4 Hippo	27.7	AD 2 Occipital Ctx (Missing)	0.0
AD 5 hippo	100.0	AD 3 Occipital Ctx	33.4
AD 6 Hippo	42.9	AD 4 Occipital Ctx	48.6
Control 2 Hippo	29.3	AD 5 Occipital Ctx	57.8
Control 4 Hippo	39.2	AD 6 Occipital Ctx	43.5
Control (Path) 3 Hippo	27.4	Control 1 Occipital Ctx	14.4
AD 1 Temporal Ctx	79.6	Control 2 Occipital Ctx	73.2
AD 2 Temporal Ctx	55.5	Control 3 Occipital Ctx	85.3
AD 3 Temporal Ctx	40.1	Control 4 Occipital Ctx	28.9
AD 4 Temporal Ctx	52.1	Control (Path) 1 Occipital Ctx	69.7
AD 5 Inf Temporal Ctx	84.7	Control (Path) 2 Occipital Ctx	49.3
AD 5 Sup Temporal Ctx	79.0	Control (Path) 3 Occipital Ctx	23.3
AD 6 Inf Temporal Ctx	51.8	Control (Path) 4 Occipital Ctx	57.0
AD 6 Sup Temporal Ctx	93.3	Control 1 Parietal Ctx	22.2
Control 1 Temporal Ctx	22.5	Control 2 Parietal Ctx	84.1
Control 2 Temporal Ctx	54.0	Control 3 Parietal Ctx	27.9
Control 3 Temporal Ctx	50.0	Control (Path) 1 Parietal Ctx	68.8
Control 4 Temporal Ctx	41.5	Control (Path) 2 Parietal Ctx	54.3

Control (Path) 1 Temporal Ctx	97.9	Control (Path) 3 Parietal Ctx	26.4
Control (Path) 2 Temporal Ctx	69.3	Control (Path) 4 Parietal Ctx	78.5

Table AOG. General\_screening\_panel\_v1.5

Tissue Name	Rel. Exp.(%) Ag813, Run 247945092	Tissue Name	Rel. Exp.(%) Ag813, Run 247945092
Adipose	15.2	Renal ca. TK-10	50.3
Melanoma* Hs688(A).T	26.2	Bladder	88.3
Melanoma* Hs688(B).T	42.6	Gastric ca. (liver met.) NCI-N87	0.0
Melanoma* M14	0.0	Gastric ca. KATO III	0.0
Melanoma* LOXIMVI	4.7	Colon ca. SW-948	0.0
Melanoma* SK- MEL-5	0.0	Colon ca. SW480	0.3
Squamous cell carcinoma SCC-4	0.0	Colon ca.* (SW480 met) SW620	0.0
Testis Pool	9.6	Colon ca. HT29	0.0
Prostate ca.* (bone met) PC-3	0.0	Colon ca. HCT-116	3.2
Prostate Pool	4.6	Colon ca. CaCo-2	0.4
Placenta	36.6	Colon cancer tissue	30.1
Uterus Pool	5.4	Colon ca. SW1116	0.0
Ovarian ca. . OVCAR-3	0.0	Colon ca. Colo-205	0.0
Ovarian ca. SK-OV- 3	1.9	Colon ca. SW-48	0.0
Ovarian ca. OVCAR-4	1.2	Colon Pool	29.7
Ovarian ca. OVCAR-5	14.3	Small Intestine Pool	9.5
Ovarian ca. IGROV-1	9.7	Stomach Pool	16.3
Ovarian ca. OVCAR-8	24.1	Bone Marrow Pool	7.9
Ovary	20.6	Fetal Heart	11.8
Breast ca. MCF-7	0.0	Heart Pool	10.9
Breast ca. MDA- MB-231	0.0	Lymph Node Pool	31.4
Breast ca. BT 549	15.7	Fetal Skeletal Muscle	15.7

Breast ca. T47D	1.0	Skeletal Muscle Pool	4.1
Breast ca. MDA-N	0.0	Spleen Pool	12.3
Breast Pool	30.1	Thymus Pool	37.1
Trachea	6.0	CNS cancer (glio/astro) U87-MG	0.0
Lung	4.9	CNS cancer (glio/astro) U-118-MG	0.7
Fetal Lung	59.5	CNS cancer (neuro;met) SK-N-AS	0.6
Lung ca. NCI-N417	0.0	CNS cancer (astro) SF- 539	15.0
Lung ca. LX-1	0.0	CNS cancer (astro) SNB-75	100.0
Lung ca. NCI-H146	0.0	CNS cancer (glio) SNB-19	10.7
Lung ca. SHP-77	1.5	CNS cancer (glio) SF- 295	14.8
Lung ca. A549	75.3	Brain (Amygdala) Pool	13.7
Lung ca. NCI-H526	0.0	Brain (cerebellum)	8.5
Lung ca. NCI-H23	30.6	Brain (fetal)	95.9
Lung ca. NCI-H460	0.3	Brain (Hippocampus) Pool	12.9
Lung ca. HOP-62	19.9	Cerebral Cortex Pool	20.0
Lung ca. NCI-H522	17.7	Brain (Substantia nigra) Pool	10.5
Liver	0.4	Brain (Thalamus) Pool	22.2
Fetal Liver	6.3	Brain (whole)	12.0
Liver ca. HepG2	0.0	Spinal Cord Pool	21.0
Kidney Pool	36.6	Adrenal Gland	19.2
Fetal Kidney	36.6	Pituitary gland Pool	1.6
Renal ca. 786-0	0.0	Salivary Gland	0.4
Renal ca. A498	55.5	Thyroid (female)	4.0
Renal ca. ACHN	0.0	Pancreatic ca. CAPAN2	1.1
Renal ca. UO-31	7.0	Pancreas Pool	45.7

Table AOH. Panel 1.2

Tissue Name	Rel. Exp.(%) Ag813, Run 118348494	Rel. Exp.(%) Ag813, Run 126741639	Tissue Name	Rel. Exp.(%) Ag813, Run 118348494	Rel. Exp.(%) Ag813, Run 126741639
Endothelial cells	0.0	0.5	Renal ca. 786- 0	0.0	0.0
Heart (Fetal)	0.0	8.5	Renal ca.	0.0	7.5

			A498		
Pancreas	9.9	27.7	Renal ca. RXF 393	0.0	4.7
Pancreatic ca. CAPAN 2	0.0	0.1	Renal ca. ACHN	0.0	0.0
Adrenal Gland	15.3	79.0	Renal ca. UO-31	0.0	2.3
Thyroid	0.2	13.8	Renal ca. TK-10	5.8	18.9
Salivary gland	1.8	15.9	Liver	5.0	27.9
Pituitary gland	9.0	16.7	Liver (fetal)	1.5	12.3
Brain (fetal)	100.0	100.0	Liver ca. (hepatoblast) HepG2	0.0	0.0
Brain (whole)	15.2	33.7	Lung	2.3	22.2
Brain (amygdala)	11.4	22.7	Lung (fetal)	9.9	46.3
Brain (cerebellum)	0.3	8.1	Lung ca. (small cell) LX-1	0.0	0.0
Brain (hippocampus)	23.2	49.7	Lung ca. (small cell) NCI-H69	0.0	0.0
Brain (thalamus)	3.1	10.4	Lung ca. (s.cell var.) SHP-77	0.0	0.5
Cerebral Cortex	14.9	59.9	Lung ca. (large cell) NCI-H460	0.0	1.1
Spinal cord	6.2	29.7	Lung ca. (non-sm. cell) A549	29.9	55.5
glio/astro U87-MG	0.0	0.0	Lung ca. (non-s.cell) NCI-H23	1.9	3.7
glio/astro U-118-MG	0.0	0.1	Lung ca. (non-s.cell) HOP-62	4.1	24.0
astrocytoma SW1783	0.0	0.0	Lung ca. (non-s.cl) NCI-H522	4.4	19.9
neuro*; met SK-N-AS	0.0	0.3	Lung ca. (squam.) SW 900	0.7	8.3
astrocytoma SF-539	0.3	7.7	Lung ca. (squam.) NCI-H596	0.0	0.0
astrocytoma	0.1	4.2	Mammary	3.4	31.2

SNB-75			gland		
glioma SNB-19	0.0	8.8	Breast ca.* (pl.ef) MCF-7	0.0	0.0
glioma U251	0.0	7.6	Breast ca.* (pl.ef) MDA- MB-231	0.0	0.0
glioma SF-295	0.0	1.8	Breast ca.* (pl. ef) T47D	0.2	6.0
Heart	7.6	36.9	Breast ca. BT- 549	0.0	3.5
Skeletal Muscle	0.7	22.5	Breast ca. MDA-N	0.0	0.0
Bone marrow	0.5	3.8	Ovary	2.7	15.7
Thymus	6.8	17.2	Ovarian ca. OVCAR-3	0.0	0.2
Spleen	1.1	9.8	Ovarian ca. OVCAR-4	0.0	1.0
Lymph node	9.1	32.3	Ovarian ca. OVCAR-5	0.2	4.0
Colorectal Tissue	0.0	1.4	Ovarian ca. OVCAR-8	2.6	28.7
Stomach	1.3	27.2	Ovarian ca. IGROV-1	0.0	2.1
Small intestine	7.4	28.9	Ovarian ca. (ascites) SK- OV-3	0.0	0.3
Colon ca. SW480	0.0	0.0	Uterus	8.1	36.1
Colon ca.* SW620 (SW480 met)	0.0	0.0	Placenta	3.5	14.8
Colon ca. HT29	0.0	0.0	Prostate	0.1	16.8
Colon ca. HCT- 116	0.0	1.1	Prostate ca.* (bone met) PC-3	0.0	0.2
Colon ca. CaCo- 2	0.0	0.3	Testis	2.7	7.2
Colon ca. Tissue (ODO3866)	0.1	4.6	Melanoma Hs688(A).T	0.9	5.4
Colon ca. HCC- 2998	0.0	0.0	Melanoma* (met) Hs688(B).T	0.8	8.1
Gastric ca.* (liver met) NCI- N87	0.0	0.0	Melanoma UACC-62	0.0	0.5

Bladder	15.8	92.0	Melanoma M14	0.0	0.0
Trachea	0.7	9.9	Melanoma LOX IMVI	0.0	0.5
Kidney	1.7	16.7	Melanoma* (met) SK-MEL-5	0.0	0.0
Kidney (fetal)	11.7	67.4			

Table AOI. Panel 4.1D

Tissue Name	Rel. Exp.(%) Ag813, Run 237369996	Tissue Name	Rel. Exp.(%) Ag813, Run 237369996
Secondary Th1 act	4.8	HUVEC IL-1beta	11.6
Secondary Th2 act	11.7	HUVEC IFN gamma	6.9
Secondary Tr1 act	5.3	HUVEC TNF alpha + IFN gamma	2.7
Secondary Th1 rest	4.9	HUVEC TNF alpha + IL4	1.3
Secondary Th2 rest	2.1	HUVEC IL-11	20.0
Secondary Tr1 rest	0.0	Lung Microvascular EC none	95.9
Primary Th1 act	0.0	Lung Microvascular EC TNFalpha + IL-1beta	39.0
Primary Th2 act	61.1	Microvascular Dermal EC none	2.8
Primary Tr1 act	25.9	Microvascular Dermal EC TNFalpha + IL-1beta	6.7
Primary Th1 rest	14.7	Bronchial epithelium TNFalpha + IL1beta	0.0
Primary Th2 rest	15.8	Small airway epithelium none	0.0
Primary Tr1 rest	4.1	Small airway epithelium TNFalpha + IL-1beta	0.0
CD45RA CD4 lymphocyte act	5.4	Coronary artery SMC rest	33.4
CD45RO CD4 lymphocyte act	22.2	Coronary artery SMC TNFalpha + IL-1beta	31.4
CD8 lymphocyte act	0.9	Astrocytes rest	20.4
Secondary CD8 lymphocyte rest	6.0	Astrocytes TNFalpha + IL-1beta	6.9
Secondary CD8 lymphocyte act	0.0	KU-812 (Basophil) rest	0.0
CD4 lymphocyte none	6.0	KU-812 (Basophil) PMA/ionomycin	11.0

2ry Th1/Th2/Tr1_anti-CD95 CH11	7.1	CCD1106 (Keratinocytes) none	0.0
LAK cells rest	11.3	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	0.0
LAK cells IL-2	0.0	Liver cirrhosis	32.5
LAK cells IL-2+IL-12	0.0	NCI-H292 none	0.0
LAK cells IL-2+IFN gamma	0.0	NCI-H292 IL-4	0.0
LAK cells IL-2+ IL-18	0.0	NCI-H292 IL-9	0.0
LAK cells PMA/ionomycin	10.7	NCI-H292 IL-13	5.4
NK Cells IL-2 rest	100.0	NCI-H292 IFN gamma	0.0
Two Way MLR 3 day	13.9	HPAEC none	5.3
Two Way MLR 5 day	0.0	HPAEC TNF alpha + IL-1 beta	4.9
Two Way MLR 7 day	0.0	Lung fibroblast none	10.7
PBMC rest	2.9	Lung fibroblast TNF alpha + IL-1 beta	9.4
PBMC PWM	0.0	Lung fibroblast IL-4	7.8
PBMC PHA-L	0.0	Lung fibroblast IL-9	6.8
Ramos (B cell) none	0.0	Lung fibroblast IL-13	0.0
Ramos (B cell) ionomycin	0.0	Lung fibroblast IFN gamma	16.5
B lymphocytes PWM	2.4	Dermal fibroblast CCD1070 rest	12.2
B lymphocytes CD40L and IL-4	2.9	Dermal fibroblast CCD1070 TNF alpha	25.5
EOL-1 dbcAMP	80.1	Dermal fibroblast CCD1070 IL-1 beta	20.3
EOL-1 dbcAMP PMA/ionomycin	0.0	Dermal fibroblast IFN gamma	9.9
Dendritic cells none	0.0	Dermal fibroblast IL-4	57.8
Dendritic cells LPS	0.0	Dermal Fibroblasts rest	13.0
Dendritic cells anti-CD40	0.0	Neutrophils TNFa+LPS	0.0
Monocytes rest	0.0	Neutrophils rest	2.5
Monocytes LPS	0.0	Colon	0.0
Macrophages rest	0.0	Lung	0.0
Macrophages LPS	0.0	Thymus	4.8
HUVEC none	0.0	Kidney	21.9
HUVEC starved	14.2		

Table AOJ. Panel 5. Islet

Tissue Name	Rel. Exp.(%) Ag813, Run 254387841	Tissue Name	Rel. Exp.(%) Ag813, Run 254387841
97457_Patient-02go_adipose	45.7	94709_Donor 2 AM - A_adipose	49.3
97476_Patient-07sk_skeletal muscle	11.7	94710_Donor 2 AM - B_adipose	15.8
97477_Patient-07ut_uterus	33.2	94711_Donor 2 AM - C_adipose	8.4
97478_Patient-07pl_placenta	11.7	94712_Donor 2 AD - A_adipose	52.9
99167_Bayer Patient 1	14.4	94713_Donor 2 AD - B_adipose	36.3
97482_Patient-08ut_uterus	45.7	94714_Donor 2 AD - C_adipose	35.6
97483_Patient-08pl_placenta	7.0	94742_Donor 3 U - A Mesenchymal Stem Cells	27.4
97486_Patient-09sk_skeletal muscle	0.0	94743_Donor 3 U - B Mesenchymal Stem Cells	33.9
97487_Patient-09ut_uterus	16.3	94730_Donor 3 AM - A_adipose	17.2
97488_Patient-09pl_placenta	13.8	94731_Donor 3 AM - B_adipose	21.2
97492_Patient-10ut_uterus	24.3	94732_Donor 3 AM - C_adipose	4.9
97493_Patient-10pl_placenta	5.1	94733_Donor 3 AD - A_adipose	100.0
97495_Patient-11go_adipose	9.7	94734_Donor 3 AD - B_adipose	40.3
97496_Patient-11sk_skeletal muscle	15.0	94735_Donor 3 AD - C_adipose	69.7
97497_Patient-11ut_uterus	43.2	77138_Liver_HepG2untreated	0.0
97498_Patient-11pl_placenta	7.9	73556_Heart_Cardiac stromal cells (primary)	7.9
97500_Patient-12go_adipose	36.3	81735_Small Intestine	54.3
97501_Patient-12sk_skeletal muscle	33.2	72409_Kidney Proximal Convoluted Tubule	0.0
97502_Patient-12ut_uterus	55.1	82685_Small intestine_Duodenum	0.0
97503_Patient-12pl_placenta	0.0	90650_Adrenal_Adrenocortical adenoma	26.2
94721_Donor 2 U - A Mesenchymal	66.0	72410_Kidney_HRCE	0.0

Stem Cells			
94722_Donor 2 U - B_Mesenchymal Stem Cells	32.1	72411_Kidney_HRE	0.0
94723_Donor 2 U - C_Mesenchymal Stem Cells	62.0	73139_Uterus_Uterine smooth muscle cells	6.3

Table AOK. Panel CNS\_1

Tissue Name	Rel. Exp.(%) Ag813, Run 171629144	Tissue Name	Rel. Exp.(%) Ag813, Run 171629144
BA4 Control	3.0	BA17 PSP	3.9
BA4 Control2	22.7	BA17 PSP2	4.8
BA4 Alzheimer's2	6.6	Sub Nigra Control	14.8
BA4 Parkinson's	36.6	Sub Nigra Control2	16.5
BA4 Parkinson's2	49.3	Sub Nigra Alzheimer's2	6.1
BA4 Huntington's	8.4	Sub Nigra Parkinson's2	23.8
BA4 Huntington's2	12.0	Sub Nigra Huntington's	14.6
BA4 PSP	5.9	Sub Nigra Huntington's2	32.8
BA4 PSP2	6.7	Sub Nigra PSP2	0.0
BA4 Depression	11.0	Sub Nigra Depression	2.5
BA4 Depression2	19.6	Sub Nigra Depression2	7.6
BA7 Control	19.9	Glob Palladus Control	2.5
BA7 Control2	18.9	Glob Palladus Control2	0.7
BA7 Alzheimer's2	12.5	Glob Palladus Alzheimer's	4.6
BA7 Parkinson's	33.9	Glob Palladus Alzheimer's2	3.2
BA7 Parkinson's2	31.4	Glob Palladus Parkinson's	41.8
BA7 Huntington's	37.4	Glob Palladus Parkinson's2	11.4
BA7 Huntington's2	100.0	Glob Palladus PSP	5.0
BA7 PSP	21.9	Glob Palladus PSP2	0.0

BA7 PSP2	3.8	Glob Palladus Depression	1.7
BA7 Depression	6.2	Temp Pole Control	2.9
BA9 Control	22.2	Temp Pole Control2	4.8
BA9 Control2	23.5	Temp Pole Alzheimer's	2.5
BA9 Alzheimer's	0.0	Temp Pole Alzheimer's2	12.4
BA9 Alzheimer's2	19.8	Temp Pole Parkinson's	28.9
BA9 Parkinson's	42.6	Temp Pole Parkinson's2	13.1
BA9 Parkinson's2	21.6	Temp Pole Huntington's	28.1
BA9 Huntington's	17.7	Temp Pole PSP	6.9
BA9 Huntington's2	52.5	Temp Pole PSP2	1.9
BA9 PSP	6.8	Temp Pole Depression2	18.8
BA9 PSP2	2.7	Cing Gyr. Control	39.0
BA9 Depression	6.7	Cing Gyr. Control2	16.4
BA9 Depression2	10.4	Cing Gyr Alzheimer's	4.8
BA17 Control	43.5	Cing Gyr Alzheimer's2	7.2
BA17 Control2	24.1	Cing Gyr Parkinson's	22.4
BA17 Alzheimer's2	21.2	Cing Gyr Parkinson's2	9.2
BA17 Parkinson's	33.4	Cing Gyr Huntington's	24.3
BA17 Parkinson's2	39.0	Cing Gyr Huntington's2	33.9
BA17 Huntington's	24.0	Cing Gyr PSP	5.2
BA17 Huntington's2	37.9	Cing Gyr PSP2	0.0
BA17 Depression	31.9	Cing Gyr Depression	9.5
BA17 Depression2	45.7	Cing Gyr Depression2	0.0

**AI\_comprehensive\_panel\_v1.0 Summary:** Ag813 Two experiments with same probe-primer sets are in excellent agreement. Highest expression of this gene is detected in orthoarthritis bone (CTs=29-30.6). In addition significant expression of this gene is detected in samples derived from orthoarthritis bone, cartilage, synovium and synovial fluid  
 5 samples, from normal lung, COPD lung, emphysema, atopic asthma, asthma, allergy, Crohn's disease (normal matched control and diseased), ulcerative colitis(normal matched control and diseased), and psoriasis (normal matched control and diseased). Interestingly, expression of this gene in normal and rheumatoid arthritis bone, synovium and synovial fluid is very low or undetectable. Therefore, therapeutic modulation of this gene product  
 10 may ameliorate symptoms/conditions associated with autoimmune and inflammatory disorders including psoriasis, allergy, asthma, inflammatory bowel disease, and osteoarthritis.

**CNS\_neurodegeneration\_v1.0 Summary:** Ag813 This panel confirms the expression of this gene at low levels in the brains of an independent group of individuals. However, no  
 15 differential expression of this gene was detected between Alzheimer's diseased postmortem brains and those of non-demented controls in this experiment. Please see Panel 1.5 for a discussion of the potential utility of this gene in treatment of central nervous system disorders.

**General\_screening\_panel\_v1.5 Summary:** Ag813 Highest expression of this gene is  
 20 detected in fetal brain and brain cancer SNB-75 cell line (CTs=31). In addition, moderate expression of this gene is seen all regions of the central nervous system examined, including amygdala, hippocampus, substantia nigra, thalamus, cerebellum, cerebral cortex, and spinal cord. This gene codes for a variant of ADAMTS-10, a member of Matrix metalloproteinases (MMPs). MMPs are a gene family of neutral proteases that are  
 25 important in normal development, wound healing, and a wide variety of pathological processes, including the spread of metastatic cancer cells, arthritic destruction of joints, atherosclerosis, and neuroinflammation. In the central nervous system (CNS), MMPs have been shown to degrade components of the basal lamina, leading to disruption of the blood-brain barrier (BBB), and to contribute to the neuroinflammatory response in many  
 30 neurological diseases (Rosenberg GA, 2002, Glia 39(3):279-91, PMID: 12203394). Therefore, therapeutic modulation of this gene product may be useful in the treatment of neurological disorders such as Alzheimer's disease, Parkinson's disease, epilepsy, multiple

sclerosis, schizophrenia, depression, allergic encephalomyelitis (EAE), allergic neuritis (EAN), and cerebral ischemia.

Moderate to low expression of this gene is also detected in tissues with metabolic/endocrine function including pancreas, adipose, adrenal gland, skeletal muscle, heart, fetal liver and the gastrointestinal tract. Therefore, therapeutic modulation of the activity of this gene may prove useful in the treatment of endocrine/metabolically related diseases, such as obesity and diabetes.

In addition, this gene is expressed at moderate to low levels in number of cancer cell lines derived from melanoma, ovarian, breast, lung, renal, colon and brain cancers. Therefore, therapeutic modulation of this gene through the use of protein therapeutics, antibodies or small molecule drug may be useful in the treatment of these cancer.

Using Curagen PathCalling technology, the ADAMTS-10 protein encoded by this gene was shown to interact with amphiregulin (AREG). AREG is shown to inhibit growth of certain human tumor cells and stimulates proliferation of human fibroblasts and other normal and tumor cells (Shoyab et al., 1988, Proc. Nat. Acad. Sci. 85: 6528-6532. PubMed ID : 3413110). Recently, AREG has been implicated in the regulation of neural stem cell proliferation and neurogenesis in the adult brain.

**Panel 1.2 Summary:** Ag813 Highest expression of this gene is detected in fetal brain (CT=27.5). In addition, moderate expression of this gene is all regions of the central nervous system examined, including amygdala, hippocampus, substantia nigra, thalamus, cerebellum, cerebral cortex, and spinal cord. Moderate to low expression of this gene is also detected in tissues with metabolic/endocrine function and number of cancer cell lines derived from melanoma, ovarian, lung, renal, colon and brain cancers. Please see panel 1.5 for further discussion on the utility of this gene.

**Panel 4.1D Summary:** Ag813 Highest expression of this gene is detected in IL-2 treated resting NK cells (CT=32.8). Moderate to low levels of expression of this gene is also detected in activated primary polarized T cells, eosinophils, lung microvascular endothelial cells, coronary artery SMC, liver cirrhosis and activated dermal fibroblasts. Therefore, therapeutic modulation of this gene or the protein encoded by this gene may be useful in the treatment of autoimmune and inflammatory diseases including asthma, allergies,

inflammatory bowel disease, lupus erythematosus, psoriasis, rheumatoid arthritis, and osteoarthritis.

Results from one experiment (Run 247683477) with this gene are not included. The amp plot indicates that there were experimental difficulties with this run.

- 5 **Panel 5 Islet Summary:** Ag813 Highest expression of this gene is detected in differentiated adipose (CT=33.5). Low expression of this gene is seen mainly in adipose and small intestine. Therefore, therapeutic modulation of this gene or its protein product may be useful in the treatment of obesity and diabetes, including Type II diabetes.

- 10 **Panel CNS\_1 Summary:** Ag813 This panel confirms the expression of this gene at low levels in the brains of an independent group of individuals. Please see Panel 1.5 for a discussion of the potential utility of this gene in treatment of central nervous system disorders.

**AP. CG56155-02: PLASMA KALLIKREIN PRECURSOR.**

- 15 Expression of gene CG56155-02 was assessed using the primer-probe set Ag1688, described in Table APA. Results of the RTQ-PCR runs are shown in Tables APB, APC, APD, APE, APF, APG and APH.

Table APA. Probe Name Ag1688

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-tcagaaggaatcatgatatcg-3'	22	577	606
Probe	TET-5'-ccttgataaaactccaggctcctttga-3'-TAMRA	27	550	607
Reverse	5'-tttggaaggttaggcatttgg-3'	21	509	608

Table APB. AI\_comprehensive panel\_v1.0

Tissue Name	Rel. Exp.(%) Ag1688, Run 248429492	Tissue Name	Rel. Exp.(%) Ag1688, Run 248429492
110967 COPD-F	53.6	112427 Match Control Psoriasis-F	77.4
110980 COPD-F	14.2	112418 Psoriasis-M	12.7
110968 COPD-M	48.3	112723 Match Control Psoriasis-M	0.0

110977 COPD-M	53.6	112419 Psoriasis-M	100.0
110989 Emphysema-F	61.6	112424 Match Control Psoriasis-M	35.6
110992 Emphysema-F	21.6	112420 Psoriasis-M	87.7
110993 Emphysema-F	23.8	112425 Match Control Psoriasis-M	29.1
110994 Emphysema-F	20.7	104689 (MF) OA Bone-Backus	50.0
110995 Emphysema-F	55.1	104690 (MF) Adj "Normal" Bone-Backus	34.9
110996 Emphysema-F	17.8	104691 (MF) OA Synovium-Backus	25.5
110997 Asthma-M	25.2	104692 (BA) OA Cartilage-Backus	37.6
111001 Asthma-F	23.0	104694 (BA) OA Bone-Backus	8.4
111002 Asthma-F	22.1	104695 (BA) Adj "Normal" Bone-Backus	34.4
111003 Atopic Asthma-F	15.5	104696 (BA) OA Synovium-Backus	6.9
111004 Atopic Asthma-F	19.9	104700 (SS) OA Bone-Backus	22.8
111005 Atopic Asthma-F	23.8	104701 (SS) Adj "Normal" Bone-Backus	42.3
111006 Atopic Asthma-F	6.0	104702 (SS) OA Synovium-Backus	29.5
111417 Allergy-M	4.6	117093 OA Cartilage Rep7	10.6
112347 Allergy-M	0.0	112672 OA Bone5	94.0
112349 Normal Lung-F	0.0	112673 OA Synovium5	43.2
112357 Normal Lung-F	38.7	112674 OA Synovial Fluid cells5	58.6
112354 Normal Lung-M	24.0	117100 OA Cartilage Rep14	0.0
112374 Crohns-F	10.4	112756 OA Bone9	2.6
112389 Match Control Crohns-F	7.4	112757 OA Synovium9	8.0
112375 Crohns-F	4.6	112758 OA Synovial Fluid Cells9	22.1
112732 Match	25.0	117125 RA Cartilage	22.1

Control Crohns-F		Rep2	
112725 Crohns-M	11.3	113492 Bone2 RA	10.0
112387 Match Control Crohns-M	1.0	113493 Synovium2 RA	11.0
112378 Crohns-M	0.0	113494 Syn Fluid Cells RA	31.6
112390 Match Control Crohns-M	44.1	113499 Cartilage4 RA	47.6
112726 Crohns-M	19.5	113500 Bone4 RA	37.9
112731 Match Control Crohns-M	58.2	113501 Synovium4 RA	55.5
112380 Ulcer Col-F	3.2	113502 Syn Fluid Cells4 RA	10.0
112734 Match Control Ulcer Col-F	56.6	113495 Cartilage3 RA	20.7
112384 Ulcer Col-F	10.1	113496 Bone3 RA	16.2
112737 Match Control Ulcer Col-F	21.6	113497 Synovium3 RA	11.5
112386 Ulcer Col-F	0.0	113498 Syn Fluid Cells3 RA	25.3
112738 Match Control Ulcer Col-F	9.3	117106 Normal Cartilage Rep20	0.0
112381 Ulcer Col-M	0.0	113663 Bone3 Normal	0.9
112735 Match Control Ulcer Col-M	41.8	113664 Synovium3 Normal	0.0
112382 Ulcer Col-M	3.8	113665 Syn Fluid Cells3 Normal	1.1
112394 Match Control Ulcer Col-M	5.2	117107 Normal Cartilage Rep22	2.7
112383 Ulcer Col-M	31.6	113667 Bone4 Normal	8.1
112736 Match Control Ulcer Col-M	12.9	113668 Synovium4 Normal	5.8
112423 Psoriasis-F	9.2	113669 Syn Fluid Cells4 Normal	5.3

Table APC. CNS\_neurodegeneration\_v1.0

Tissue Name	Rel. Exp.(%) Ag1688, Run 269217573	Tissue Name	Rel. Exp.(%) Ag1688, Run 269217573
AD 1 Hippo	24.5	Control (Path) 3 Temporal Ctx	9.7
AD 2 Hippo	34.4	Control (Path) 4 Temporal Ctx	41.8
AD 3 Hippo	17.9	AD 1 Occipital	42.3

		Ctx	
AD 4 Hippo.	18.0	AD 2 Occipital Ctx (Missing)	0.0
AD 5 hippo	94.6	AD 3 Occipital Ctx	7.1
AD 6 Hippo	34.9	AD 4 Occipital Ctx	26.4
Control 2 Hippo	35.4	AD 5 Occipital Ctx	9.9
Control 4 Hippo	50.7	AD 6 Occipital Ctx	27.2
Control (Path) 3 Hippo	9.3	Control 1 Occipital Ctx	6.3
AD 1 Temporal Ctx	31.9	Control 2 Occipital Ctx	49.7
AD 2 Temporal Ctx	31.4	Control 3 Occipital Ctx	39.2
AD 3 Temporal Ctx	20.4	Control 4 Occipital Ctx	26.6
AD 4 Temporal Ctx	29.5	Control (Path) 1 Occipital Ctx	47.3
AD 5 Inf Temporal Ctx	100.0	Control (Path) 2 Occipital Ctx	21.3
AD 5 SupTemporal Ctx	92.0	Control (Path) 3 Occipital Ctx	3.5
AD 6 Inf Temporal Ctx	43.8	Control (Path) 4 Occipital Ctx	17.8
AD 6 Sup Temporal Ctx	69.7	Control 1 Parietal Ctx	19.5
Control 1 Temporal Ctx	16.5	Control 2 Parietal Ctx	85.3
Control 2 Temporal Ctx	34.9	Control 3 Parietal Ctx	15.5
Control 3 Temporal Ctx	32.3	Control (Path) 1 Parietal Ctx	44.4
Control 4 Temporal Ctx	35.4	Control (Path) 2 Parietal Ctx	52.9
Control (Path) 1 Temporal Ctx	46.0	Control (Path) 3 Parietal Ctx	9.7
Control (Path) 2 Temporal Ctx	45.7	Control (Path) 4 Parietal Ctx	52.1

Table APD. Panel 1.3D.

Tissue Name	Rel. Exp.(%)	Tissue Name	Rel. Exp.(%)
-------------	--------------	-------------	--------------

	Ag1688, Run 147249266		Ag1688, Run 147249266
Liver adenocarcinoma	0.0	Kidney (fetal)	9.2
Pancreas	6.7	Renal ca. 786-0	0.0
Pancreatic ca. CAPAN 2	0.2	Renal ca. A498	1.7
Adrenal gland	1.8	Renal ca. RXF 393	0.0
Thyroid	3.8	Renal ca. ACHN	0.0
Salivary gland	1.5	Renal ca. UO-31	0.0
Pituitary gland	6.1	Renal ca. TK-10	0.0
Brain (fetal)	0.5	Liver	100.0
Brain (whole)	3.6	Liver (fetal)	99.3
Brain (amygdala)	3.3	Liver ca. (hepatoblast) HepG2	0.0
Brain (cerebellum)	0.4	Lung	1.3
Brain (hippocampus)	6.2	Lung (fetal)	1.8
Brain (substantia nigra)	1.0	Lung ca. (small cell) LX-1	0.0
Brain (thalamus)	2.1	Lung ca. (small cell) NCI-H69	0.0
Cerebral Cortex	6.3	Lung ca. (s.cell var.) SHP-77	0.8
Spinal cord	3.1	Lung ca. (large cell)NCI-H460	0.0
glio/astro U87-MG	0.0	Lung ca. (non-sm. cell) A549	0.2
glio/astro U-118-MG	0.0	Lung ca. (non-s.cell) NCI-H23	0.0
astrocytoma SW1783	0.0	Lung ca. (non-s.cell) HOP-62	0.0
neuro*; met SK-N-AS	0.2	Lung ca. (non-s.cl) NCI-H522	0.0
astrocytoma SF-539	0.0	Lung ca. (squam.) SW 900	0.2
astrocytoma SNB-75	0.1	Lung ca. (squam.) NCI-H596	0.0
glioma SNB-19	0.2	Mammary gland	2.9
glioma U251	1.2	Breast ca.* (pl.ef) MCF-7	0.0
glioma SF-295	0.0	Breast ca.* (pl.ef) MDA-MB-231	0.0
Heart (fetal)	0.2	Breast ca.* (pl.ef) T47D	0.0
Heart	1.6	Breast ca. BT-549	0.0

Skeletal muscle (fetal)	0.7	Breast ca. MDA-N	0.0
Skeletal muscle	1.2	Ovary	0.0
Bone marrow	0.5	Ovarian ca. OVCAR-3	0.2
Thymus	3.2	Ovarian ca. OVCAR-4	0.0
Spleen	1.0	Ovarian ca. OVCAR-5	0.3
Lymph node	2.9	Ovarian ca. OVCAR-8	0.0
Colorectal	0.8	Ovarian ca. IGROV-1	0.0
Stomach	3.3	Ovarian ca.* (ascites) SK-OV-3	1.0
Small intestine	6.2	Uterus	1.4
Colon ca. SW480	0.0	Placenta	0.4
Colon ca.* SW620(SW480 met)	0.0	Prostate	1.0
Colon ca. HT29	0.0	Prostate ca.* (bone met)PC-3	0.0
Colon ca. HCT-116	0.0	Testis	6.1
Colon ca. CaCo-2	0.2	Melanoma Hs688(A).T	0.4
Colon ca. tissue(ODO3866)	0.0	Melanoma* (met) Hs688(B).T	0.9
Colon ca. HCC-2998	0.2	Melanoma UACC-62	0.0
Gastric ca.* (liver met) NCI-N87	4.4	Melanoma M14	0.0
Bladder	3.1	Melanoma LOX IMVI	0.0
Trachea	3.0	Melanoma* (met) SK-MEL-5	0.0
Kidney	6.8	Adipose	0.5

Table APE. Panel 2D

Tissue Name	Rel. Exp.(%) Ag1688, Run 162646059	Tissue Name	Rel. Exp.(%) Ag1688, Run 162646059
Normal Colon	1.7	Kidney Margin 8120608	0.7
CC Well to Mod Diff (ODO3866)	0.0	Kidney Cancer 8120613	0.0

CC Margin (ODO3866)	0.2	Kidney Margin 8120614	0.5
CC Gr.2 rectosigmoid (ODO3868)	0.2	Kidney Cancer 9010320	0.2
CC Margin (ODO3868)	0.1	Kidney Margin 9010321	1.0
CC Mod Diff (ODO3920)	0.1	Normal Uterus	0.2
CC Margin (ODO3920)	0.9	Uterus Cancer 064011	0.8
CC Gr.2 ascend colon (ODO3921)	0.1	Normal Thyroid	0.9
CC Margin (ODO3921)	0.1	Thyroid Cancer 064010	0.2
CC from Partial Hepatectomy (ODO4309) Mets	4.7	Thyroid Cancer A302152	0.5
Liver Margin (ODO4309)	100.0	Thyroid Margin A302153	1.0
Colon mets to lung (OD04451-01)	0.1	Normal Breast	0.3
Lung Margin (OD04451- 02)	0.1	Breast Cancer (OD04566)	0.1
Normal Prostate 6546-1	2.1	Breast Cancer (OD04590-01)	0.1
Prostate Cancer (OD04410)	0.6	Breast Cancer Mets (OD04590-03)	0.4
Prostate Margin (OD04410)	0.5	Breast Cancer Metastasis (OD04655-05)	0.9
Prostate Cancer (OD04720-01)	1.1	Breast Cancer 064006	0.6
Prostate Margin (OD04720-02)	1.6	Breast Cancer 1024	1.2
Normal Lung 061010	2.0	Breast Cancer 9100266	0.1
Lung Met to Muscle (ODO4286)	0.0	Breast Margin 9100265	0.1
Muscle Margin (ODO4286)	0.2	Breast Cancer A209073	0.3
Lung Malignant Cancer (OD03126)	0.1	Breast Margin A209073	0.3
Lung Margin (OD03126)	0.5	Normal Liver	69.7
Lung Cancer (OD04404)	0.1	Liver Cancer 064003	13.7
Lung Margin (OD04404)	0.2	Liver Cancer 1025	18.0

Lung Cancer (OD04565)	0.0	Liver Cancer 1026	1.2
Lung Margin (OD04565)	0.1	Liver Cancer 6004-T	22.2
Lung Cancer (OD04237-01)	0.1	Liver Tissue 6004-N	1.0
Lung Margin (OD04237-02)	0.4	Liver Cancer 6005-T	1.9
Ocular Mel Met to Liver (ODO4310)	0.1	Liver Tissue 6005-N	4.2
Liver Margin (ODO4310)	77.4	Normal Bladder	2.7
Melanoma Mets to Lung (OD04321)	0.0	Bladder Cancer 1023	0.0
Lung Margin (OD04321)	0.1	Bladder Cancer A302173	0.2
Normal Kidney	12.9	Bladder Cancer (OD04718-01)	0.1
Kidney Ca, Nuclear grade 2 (OD04338)	3.8	Bladder Normal Adjacent (OD04718-03)	0.5
Kidney Margin (OD04338)	1.6	Normal Ovary	0.0
Kidney Ca Nuclear grade 1/2 (OD04339)	2.8	Ovarian Cancer 064008	0.1
Kidney Margin (OD04339)	9.3	Ovarian Cancer (OD04768-07)	0.2
Kidney Ca, Clear cell type (OD04340)	1.4	Ovary Margin (OD04768-08)	0.1
Kidney Margin (OD04340)	4.1	Normal Stomach	0.3
Kidney Ca, Nuclear grade 3 (OD04348)	0.1	Gastric Cancer 9060358	0.1
Kidney Margin (OD04348)	3.8	Stomach Margin 9060359	0.0
Kidney Cancer (OD04622-01)	0.2	Gastric Cancer 9060395	0.2
Kidney Margin (OD04622-03)	0.7	Stomach Margin 9060394	0.3
Kidney Cancer (OD04450-01)	0.2	Gastric Cancer 9060397	0.3
Kidney Margin (OD04450-03)	2.6	Stomach Margin 9060396	0.0
Kidney Cancer 8120607	0.0	Gastric Cancer 064005	1.1

Table APF. Panel 4.1D

Tissue Name	Rel. Exp.(%) Ag1688, Run 248389308	Tissue Name	Rel. Exp.(%) Ag1688, Run 248389308
Secondary Th1 act	1.6	HUVEC IL-1beta	0.0
Secondary Th2 act	1.7	HUVEC IFN gamma	0.0
Secondary Tr1 act	0.0	HUVEC TNF alpha + IFN gamma	0.0
Secondary Th1 rest	0.0	HUVEC TNF alpha + IL4	0.0
Secondary Th2 rest	0.0	HUVEC IL-11	0.0
Secondary Tr1 rest	0.0	Lung Microvascular EC none	0.0
Primary Th1 act	0.0	Lung Microvascular EC TNFalpha + IL-1beta	0.0
Primary Th2 act	0.0	Microvascular Dermal EC none	0.0
Primary Tr1 act	0.0	Microvascular Dermal EC TNFalpha + IL-1beta	0.0
Primary Th1 rest	1.3	Bronchial epithelium TNFalpha + IL1beta	0.0
Primary Th2 rest	1.3	Small airway epithelium none	0.0
Primary Tr1 rest	1.6	Small airway epithelium TNFalpha + IL-1beta	0.0
CD45RA CD4 lymphocyte act	3.5	Coronary artery SMC rest	0.0
CD45RO CD4 lymphocyte act	4.2	Coronary artery SMC TNFalpha + IL-1beta	0.0
CD8 lymphocyte act	3.2	Astrocytes rest	0.0
Secondary CD8 lymphocyte rest	1.8	Astrocytes TNFalpha + IL-1beta	2.4
Secondary CD8 lymphocyte act	0.0	KU-812 (Basophil) rest	1.8
CD4 lymphocyte none	3.8	KU-812 (Basophil) PMA/ionomycin	0.0
2ry Th1/Th2/Tr1_anti- CD95 CH11	0.0	CCD1106 (Keratinocytes) none	0.0
LAK cells rest	0.0	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	0.0
LAK cells IL-2	6.2	Liver cirrhosis	100.0
LAK cells IL-2+IL-12	0.0	NCI-H292 none	0.0
LAK cells IL-2+IFN gamma	1.7	NCI-H292 IL-4	1.5
LAK cells IL-2+ IL-18	3.4	NCI-H292 IL-9	1.9

LAK cells PMA/ionomycin	0.0	NCI-H292 IL-13	0.0
NK Cells IL-2 rest	22.1	NCI-H292 IFN gamma	0.0
Two Way MLR 3 day	3.3	HPAEC none	0.0
Two Way MLR 5 day	1.9	HPAEC TNF alpha + IL-1 beta	0.0
Two Way MLR 7 day	1.7	Lung fibroblast none	2.6
PBMC rest	1.5	Lung fibroblast TNF alpha + IL-1 beta	10.4
PBMC PWM	5.1	Lung fibroblast IL-4	1.8
PBMC PHA-L	0.7	Lung fibroblast IL-9	12.3
Ramos (B cell) none	0.0	Lung fibroblast IL-13	0.0
Ramos (B cell) ionomycin	0.0	Lung fibroblast IFN gamma	3.1
B lymphocytes PWM	2.8	Dermal fibroblast CCD1070 rest	0.0
B lymphocytes CD40L and IL-4	21.5	Dermal fibroblast CCD1070 TNF alpha	0.0
EOL-1 dbcAMP	0.0	Dermal fibroblast CCD1070 IL-1 beta	0.0
EOL-1 dbcAMP PMA/ionomycin	0.0	Dermal fibroblast IFN gamma	6.8
Dendritic cells none	2.0	Dermal fibroblast IL-4	5.8
Dendritic cells LPS	0.0	Dermal Fibroblasts rest	0.0
Dendritic cells anti-CD40	4.9	Neutrophils TNFa+LPS	0.0
Monocytes rest	0.0	Neutrophils rest	0.0
Monocytes LPS	0.0	Colon	0.0
Macrophages rest	0.0	Lung	1.2
Macrophages LPS	0.0	Thymus	0.0
HUVEC none	0.0	Kidney	82.9
HUVEC starved	0.0		

Table APG. Panel 5 Islet

Tissue Name	Rel. Exp.(%) Ag1688, Run 226587524	Tissue Name	Rel. Exp.(%) Ag1688, Run 226587524
97457_Patient-02go_adipose	41.2	94709_Donor 2 AM - A_adipose	0.0
97476_Patient-07sk_skeletal muscle	9.9	94710_Donor 2 AM - B_adipose	0.0
97477_Patient-07ut_uterus	8.1	94711_Donor 2 AM - C_adipose	0.0

97478_Patient-07pl_placenta	0.0	94712_Donor 2 AD - A_adipose	11.4
99167_Bayer Patient 1	84.7	94713_Donor 2 AD - B_adipose	0.0
97482_Patient-08ut_uterus	2.4	94714_Donor 2 AD - C_adipose	29.1
97483_Patient-08pl_placenta	0.0	94742_Donor 3 U - A_Mesenchymal Stem Cells	19.2
97486_Patient-09sk_skeletal muscle	8.0	94743_Donor 3 U - B_Mesenchymal Stem Cells	0.0
97487_Patient-09ut_uterus	9.6	94730_Donor 3 AM - A_adipose	15.0
97488_Patient-09pl_placenta	0.0	94731_Donor 3 AM - B_adipose	37.9
97492_Patient-10ut_uterus	0.0	94732_Donor 3 AM - C_adipose	0.0
97493_Patient-10pl_placenta	0.0	94733_Donor 3 AD - A_adipose	39.2
97495_Patient-11go_adipose	0.0	94734_Donor 3 AD - B_adipose	11.4
97496_Patient-11sk_skeletal muscle	52.9	94735_Donor 3 AD - C_adipose	34.4
97497_Patient-11ut_uterus	35.8	77138_Liver_HepG2untreated	8.4
97498_Patient-11pl_placenta	10.5	73556_Heart_Cardiac stromal cells (primary)	0.0
97500_Patient-12go_adipose	0.0	81735_Small Intestine	100.0
97501_Patient-12sk_skeletal muscle	35.4	72409_Kidney_Proximal Convoluted Tubule	9.9
97502_Patient-12ut_uterus	20.7	82685_Small intestine_Duodenum	70.2
97503_Patient-12pl_placenta	0.0	90650_Adrenal_Adrenocortical adenoma	25.5
94721_Donor 2 U - A_Mesenchymal Stem Cells	0.0	72410_Kidney_HRCE	10.4
94722_Donor 2 U - B_Mesenchymal Stem Cells	0.0	72411_Kidney_HRE	7.2
94723_Donor 2 U - C_Mesenchymal Stem Cells	0.0	73139_Uterus_Uterine smooth muscle cells	0.0

Table APH. general oncology screening panel\_v\_2.4

Tissue Name	Rel. Exp.(%) Ag1688, Run 260552690	Tissue Name	Rel. Exp.(%) Ag1688, Run 260552690
Colon cancer 1	1.8	Bladder cancer NAT 2	0.1
Colon cancer NAT 1	1.0	Bladder cancer NAT 3	0.0
Colon cancer 2	0.4	Bladder cancer NAT 4	1.1
Colon cancer NAT 2	1.2	Prostate adenocarcinoma 1	3.7
Colon cancer 3	0.8	Prostate adenocarcinoma 2	0.2
Colon cancer NAT 3	2.5	Prostate adenocarcinoma 3	1.2
Colon malignant cancer 4	2.1	Prostate adenocarcinoma 4	3.5
Colon normal adjacent tissue 4	0.2	Prostate cancer NAT 5	0.6
Lung cancer 1	0.2	Prostate adenocarcinoma 6	0.2
Lung NAT 1	0.2	Prostate adenocarcinoma 7	0.0
Lung cancer 2	1.0	Prostate adenocarcinoma 8	0.0
Lung NAT 2	0.8	Prostate adenocarcinoma 9	0.0
Squamous cell carcinoma 3	0.5	Prostate cancer NAT 10	0.1
Lung NAT 3	0.0	Kidney cancer 1	7.7
metastatic melanoma 1	1.1	Kidney NAT 1	5.7
Melanoma 2	0.1	Kidney cancer 2	40.1
Melanoma 3	0.0	Kidney NAT 2	23.8
metastatic melanoma 4	2.0	Kidney cancer 3	100.0
metastatic melanoma 5	3.0	Kidney NAT 3	5.6
Bladder cancer 1	0.6	Kidney cancer 4	2.0
Bladder cancer NAT 1	0.0	Kidney NAT 4	4.2
Bladder cancer 2	0.3		

**AI\_comprehensive panel\_v1.0 Summary:** Ag1688 Highest expression of this gene is detected in psoriasis sample (CT=31.9). Moderate to low levels of expression of this gene

is also seen in samples derived from orthoarthritis/ rheumatoid arthritis bone, cartilage, synovium and synovial fluid samples, from normal lung, COPD lung, emphysema, atopic asthma, asthma, Crohn's disease (normal matched control and diseased), ulcerative colitis(normal matched control and diseased), and psoriasis (normal matched control and diseased). Therefore, therapeutic modulation of this gene product may ameliorate symptoms/conditions associated with autoimmune and inflammatory disorders including psoriasis, asthma, inflammatory bowel disease, rheumatoid arthritis and osteoarthritis.

**CNS\_neurodegeneration\_v1.0 Summary:** Ag1688 This panel confirms the expression of this gene at low levels in the brains of an independent group of individuals. However, no differential expression of this gene was detected between Alzheimer's diseased postmortem brains and those of non-demented controls in this experiment. Please see Panel 1.3D for a discussion of the potential utility of this gene in treatment of central nervous system disorders.

**Panel 1.3D Summary:** Ag1688 Expression of this gene, a plasma kallikrein, is significantly higher in liver (CTs=28) than in any other sample on this panel. Thus, expression of this gene could be used as a marker of liver tissue. In addition, low levels of expression of this gene is also detected in tissues with metabolic/endocrine functions including pancreas, adrenal gland, thyroid, pituitary gland, skeletal muscle, heart, and the gastrointestinal tract. Plasma prekallikrein is a glycoprotein that participates in the surface-dependent activation of blood coagulation, fibrinolysis, kinin generation and inflammation. It is synthesized in the liver and secreted into the blood as a single polypeptide chain. It is converted to plasma kallikrein by factor XIIa. Recently, plasma kallikrein has been implicated in adipose differentiation by remodeling of the fibronectin-rich ECM of preadipocytes. Plg -/- mice show a reduction of fat deposit (Ref. 1, 2). At Curagen, it was found that plasma kallikrein significantly down-regulated in the liver of mice with 'lean' phenotype. Thus, based on Curagen GeneCalling data it is hypothesized that plasma kallikrein might cause disruption of adipose differentiation thus leading to obesity if over expressed and to a leaner phenotype if expression is below normal. Therefore, an antagonist to this gene product in the form of small molecule or antibody may be beneficial in the treatment of obesity.

Moderate to low levels of expression of this gene is also seen levels in some of the regions of central nervous system examined, including amygdala, hippocampus, substantia nigra, thalamus, cerebral cortex, and spinal cord. Therefore, therapeutic modulation of this gene product may be useful in the treatment of central nervous system disorders such as

- 5 Alzheimer's disease, Parkinson's disease, epilepsy, multiple sclerosis, schizophrenia and depression.

#### References:

1. Hoover-Plow J, Yuen L. Plasminogen binding is increased with adipocyte differentiation. *Biochem.Biophys.Res.Commun.* (2001) 284, 389-394. PMID: 11394891.
- 10 2. Selvarajan S, Lund LR, Takeuchi T, Craik CS, Werb Z. A plasma kallikrein-dependent plasminogen cascade required for adipocyte differentiation. *Nature Cell Biol.* (2001) 3, 267-275. PMID: 11231576

- Panel 2D Summary:** Ag1688 The expression of the CG56155-01 gene appears to be highest in a sample derived from a sample of normal liver tissue adjacent to a metastatic colon cancer CT=26.2). In addition, there is substantial expression in other samples of normal liver, and to a much lesser degree, malignant liver tissue. This liver specific expression is consistent with the expression seen in Panel 1.3D. Thus, the expression of this gene could be used to distinguish liver derived tissue from the other samples in the panel, and more specifically the expression of this gene could be used to distinguish normal liver from malignant liver tissue. Moreover, therapeutic modulation of this gene, through the use of small molecule drugs, protein therapeutics or antibodies might be of benefit in the treatment of liver cancer.
- 15
- 20

- Panel 4.1D Summary:** Ag1688 Highest expression of this gene is detected in liver cirrhosis (CT=31.8). In addition, moderate to low levels of expression of this gene in IL-2 treated NK cells, CD40L and IL-4 treated B lymphocytes and normal kidney. Therefore, therapeutic modulation of the protein encoded for by this gene may be useful in the treatment of inflammatory or autoimmune diseases which effect the liver and kidney including liver cirrhosis and fibrosis, lupus erythematosus and glomerulonephritis.
- 25

**Panel 5 Islet Summary:** Ag1688 Expression of the CG56155-01 gene is limited to pancreatic islets and small intestines. Please see Panel 1.3 for discussion of utility of this gene in metabolic disease.

- 5 **General oncology screening panel\_v\_2.4 Summary:** Ag1688 Highest expression of this gene is detected in kidney cancer (CT=28.4). Higher expression of this gene is associated with cancer compared to normal kidney. Therefore, expression of this gene may be used as diagnostic marker for kidney cancer and therapeutic modulation of this gene or protein encoded by this gene may through the use of antibodies or small molecule drug may be useful in the treatment of kidney cancer.

10 **AQ. CG59595-01: Ribonuclease 6 precursor.**

Expression of gene CG59595-01 was assessed using the primer-probe set Ag3488, described in Table AQA. Results of the RTQ-PCR runs are shown in Tables AQB, AQC, AQD, AQE, AQF and AQG.

Table AQA. Probe Name Ag3488

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-aactgtgcctcactaagcaaga-3'	22	963	609
Probe	TET-5'-agcagctgcaaaactgcaccgag-3'-TAMRA	23	987	610
Reverse	5'-catttgccagccagacttc-3'	19	1037	611

15 Table AQB. CNS\_neurodegeneration\_v1.0

Tissue Name	Rel. Exp.(%) Ag3488, Run 206533698	Tissue Name	Rel. Exp.(%) Ag3488, Run 206533698
AD 1 Hippo	54.0	Control (Path) 3. Temporal Ctx	8.8
AD 2 Hippo	72.7	Control (Path) 4. Temporal Ctx	42.6
AD 3 Hippo	34.2	AD 1 Occipital Ctx	36.3
AD 4 Hippo	34.4	AD 2 Occipital Ctx (Missing)	0.0
AD 5 hippo	74.7	AD 3 Occipital Ctx	20.7
AD 6 Hippo	70.2	AD 4 Occipital Ctx	31.4

Control 2 Hippo	63.3	AD 5 Occipital Ctx	22.1
Control 4 Hippo	47.6	AD 6 Occipital Ctx	42.6
Control (Path) 3 Hippo	11.3	Control 1 Occipital Ctx	7.1
AD 1 Temporal Ctx	43.5	Control 2 Occipital Ctx	47.3
AD 2 Temporal Ctx	42.0	Control 3 Occipital Ctx	21.6
AD 3 Temporal Ctx	25.9	Control 4 Occipital Ctx	18.3
AD 4 Temporal Ctx	37.6	Control (Path) 1 Occipital Ctx	63.7
AD 5 Inf Temporal Ctx	93.3	Control (Path) 2 Occipital Ctx	15.2
AD 5 Sup Temporal Ctx	100.0	Control (Path) 3 Occipital Ctx	5.2
AD 6 Inf Temporal Ctx	74.7	Control (Path) 4 Occipital Ctx	27.4
AD 6 Sup Temporal Ctx	56.3	Control 1 Parietal Ctx	12.5
Control 1 Temporal Ctx	15.6	Control 2 Parietal Ctx	59.9
Control 2 Temporal Ctx	57.8	Control 3 Parietal Ctx	25.2
Control 3 Temporal Ctx	29.3	Control (Path) 1 Parietal Ctx	57.0
Control 4 Temporal Ctx	24.8	Control (Path) 2 Parietal Ctx	30.4
Control (Path) 1 Temporal Ctx	62.0	Control (Path) 3 Parietal Ctx	3.8
Control (Path) 2 Temporal Ctx	29.5	Control (Path) 4 Parietal Ctx	51.8

Table AQC. General\_screening\_panel\_v1.4.

Tissue Name	Rel. Exp.(%) Ag3488, Run 213390581	Tissue Name	Rel. Exp.(%) Ag3488, Run 213390581
Adipose	4.1	Renal ca. TK-10	22.8
Melanoma* Hs688(A).T	2.6	Bladder	14.8
Melanoma* Hs688(B).T	1.6	Gastric ca. (liver met.) NCI-N87	5.8

Melanoma* M14	2.1	Gastric ca. KATO III	22.2
Melanoma* LOXIMVI	0.1	Colon ca. SW-948	6.0
Melanoma* SK-MEL-5	2.1	Colon ca. SW480	6.4
Squamous cell carcinoma SCC-4	2.1	Colon ca.* (SW480 met) SW620	3.3
Testis Pool	3.3	Colon ca. HT29	17.1
Prostate ca.* (bone met) PC-3	3.1	Colon ca. HCT-116	6.3
Prostate Pool	5.3	Colon ca. CaCo-2	10.6
Placenta	2.1	Colon cancer tissue	16.2
Uterus Pool	1.7	Colon ca. SW1116	6.8
Ovarian ca. OVCAR-3	4.9	Colon ca. Colo-205	1.0
Ovarian ca. SK-OV-3	27.5	Colon ca. SW-48	6.7
Ovarian ca. OVCAR-4	10.7	Colon Pool	5.5
Ovarian ca. OVCAR-5	7.0	Small Intestine Pool	6.4
Ovarian ca. IGROV-1	57.0	Stomach Pool	3.6
Ovarian ca. OVCAR-8	1.4	Bone Marrow Pool	2.5
Ovary	3.2	Fetal Heart	1.5
Breast ca. MCF-7	15.3	Heart Pool	1.8
Breast ca. MDA-MB-231	11.8	Lymph Node Pool	6.2
Breast ca. BT 549	5.4	Fetal Skeletal Muscle	0.9
Breast ca. T47D	13.0	Skeletal Muscle Pool	0.9
Breast ca. MDA-N	1.5	Spleen Pool	10.6
Breast Pool	7.1	Thymus Pool	12.2
Trachea	7.3	CNS cancer (glio/astro) U87-MG	4.8
Lung	2.8	CNS cancer (glio/astro) U-118-MG	2.2
Fetal Lung	6.8	CNS cancer (neuro;met) SK-N-AS	1.7
Lung ca. NCI-N417	0.4	CNS cancer (astro) SF-539	0.3
Lung ca. LX-1	7.3	CNS cancer (astro) SNB-75	1.8
Lung ca. NCI-H146	1.5	CNS cancer (glio)	47.3

		SNB-19	
Lung ca. SHP-77	6.7	CNS cancer (glio) SF-295	7.9
Lung ca. A549	2.4	Brain (Amygdala) Pool	3.9
Lung ca. NCI-H526	1.5	Brain (cerebellum)	2.6
Lung ca. NCI-H23	3.6	Brain (fetal)	2.5
Lung ca. NCI-H460	3.1	Brain (Hippocampus) Pool	2.1
Lung ca. HOP-62	2.9	Cerebral Cortex Pool	2.5
Lung ca. NCI-H522	3.0	Brain (Substantia nigra) Pool	3.4
Liver	0.8	Brain (Thalamus) Pool	2.8
Fetal Liver	5.9	Brain (whole)	1.3
Liver ca. HepG2	37.6	Spinal Cord Pool	6.5
Kidney Pool	8.9	Adrenal Gland	3.1
Fetal Kidney	5.5	Pituitary gland Pool	1.8
Renal ca. 786-0	100.0	Salivary Gland	9.5
Renal ca. A498	17.9	Thyroid (female)	7.0
Renal ca. ACHN	1.8	Pancreatic ca. CAPAN2	2.6
Renal ca. UO-31	6.7	Pancreas Pool	13.3

Table AQD. Panel 2.2

Tissue Name	Rel. Exp.(%) Ag3488, Run 174285071	Tissue Name	Rel. Exp.(%) Ag3488, Run 174285071
Normal Colon	12.2	Kidney Margin (OD04348)	21.5
Colon cancer (OD06064)	8.0	Kidney malignant cancer (OD06204B)	11.8
Colon Margin (OD06064)	6.6	Kidney normal adjacent tissue (OD06204E)	4.9
Colon cancer (OD06159)	5.3	Kidney Cancer (OD04450-01)	100.0
Colon Margin (OD06159)	6.7	Kidney Margin (OD04450-03)	4.8
Colon cancer (OD06297-04)	4.9	Kidney Cancer 8120613	0.9
Colon Margin (OD06297-05)	8.5	Kidney Margin 8120614	3.1
CC Gr.2 ascend colon (ODO3921)	10.4	Kidney Cancer 9010320	23.7

CC Margin (ODO3921)	9.0	Kidney Margin 9010321	2.4
Colon cancer metastasis (OD06104)	11.0	Kidney Cancer 8120607	12.1
Lung Margin (OD06104)	8.9	Kidney Margin 8120608	3.0
Colon mets to lung (OD04451-01)	19.6	Normal Uterus	11.3
Lung Margin (OD04451-02)	9.0	Uterine Cancer 064011	16.4
Normal Prostate	11.5	Normal Thyroid	5.1
Prostate Cancer (OD04410)	4.9	Thyroid Cancer 064010	4.9
Prostate Margin (OD04410)	4.7	Thyroid Cancer A302152	8.7
Normal Ovary	7.3	Thyroid Margin A302153	6.5
Ovarian cancer (OD06283-03)	8.7	Normal Breast	9.9
Ovarian Margin (OD06283-07)	4.6	Breast Cancer (OD04566)	5.7
Ovarian Cancer 064008	13.7	Breast Cancer 1024	10.8
Ovarian cancer (OD06145)	12.2	Breast Cancer (OD04590-01)	39.8
Ovarian Margin (OD06145)	18.9	Breast Cancer Mets (OD04590-03)	8.8
Ovarian cancer (OD06455-03)	80.7	Breast Cancer Metastasis (OD04655-05)	9.2
Ovarian Margin (OD06455-07)	2.4	Breast Cancer 064006	10.0
Normal Lung	7.9	Breast Cancer 9100266	7.8
Invasive poor diff. lung adeno (ODO4945-01)	14.5	Breast Margin 9100265	5.0
Lung Margin (ODO4945-03)	8.8	Breast Cancer A209073	6.0
Lung Malignant Cancer (OD03126)	26.6	Breast Margin A2090734	10.2
Lung Margin (OD03126)	4.8	Breast cancer (OD06083)	18.6
Lung Cancer (OD05014A)	7.9	Breast cancer node metastasis (OD06083)	16.6
Lung Margin (OD05014B)	23.3	Normal Liver	8.0
Lung cancer	2.8	Liver Cancer 1026	5.0

(OD06081)			
Lung Margin (OD06081)	4.0	Liver Cancer 1025	18.4
Lung Cancer (OD04237-01)	6.0	Liver Cancer 6004-T	12.8
Lung Margin (OD04237-02)	19.6	Liver Tissue 6004-N	11.0
Ocular Melanoma Metastasis	4.6	Liver Cancer 6005-T	9.7
Ocular Melanoma Margin (Liver)	10.0	Liver Tissue 6005-N	19.9
Melanoma Metastasis	6.9	Liver Cancer 064003	11.4
Melanoma Margin (Lung)	10.2	Normal Bladder	11.6
Normal Kidney	2.9	Bladder Cancer 1023	6.1
Kidney Ca, Nuclear grade 2 (OD04338)	12.2	Bladder Cancer A302173	12.0
Kidney Margin (OD04338)	9.0	Normal Stomach	23.5
Kidney Ca Nuclear grade 1/2 (OD04339)	22.7	Gastric Cancer 9060397	3.0
Kidney Margin (OD04339)	3.3	Stomach Margin 9060396	12.7
Kidney Ca, Clear cell type (OD04340)	17.8	Gastric Cancer 9060395	8.0
Kidney Margin (OD04340)	8.0	Stomach Margin 9060394	26.4
Kidney Ca, Nuclear grade 3 (OD04348)	5.8	Gastric Cancer 064005	6.3

Table AQE. Panel 3D

Tissue Name	Rel. Exp.(%) Ag3488, Run 182098858	Tissue Name	Rel. Exp.(%) Ag3488, Run 182098858
Daoy- Medulloblastoma	1.7	Ca Ski- Cervical epidermoid carcinoma (metastasis)	18.6
TE671- Medulloblastoma	10.2	ES-2- Ovarian clear cell carcinoma	10.2
D283 Med- Medulloblastoma	34.6	Ramos- Stimulated with PMA/ionomycin 6h	7.3
PFSK-1- Primitive Neuroectodermal	11.9	Ramos- Stimulated with PMA/ionomycin 14h	27.7
XF-498- CNS	3.5	MEG-01- Chronic myelogenous leukemia	27.2

		(megokaryoblast)	
SNB-78- Glioma	21.5	Raji- Burkitt's lymphoma	16.0
SF-268- Glioblastoma	11.9	Daudi- Burkitt's lymphoma	8.8
T98G- Glioblastoma	5.3	U266- B-cell plasmacytoma	17.3
SK-N-SH- Neuroblastoma (metastasis)	22.5	CA46- Burkitt's lymphoma	6.4
SF-295- Glioblastoma	10.4	RL- non-Hodgkin's B-cell lymphoma	2.9
Cerebellum	11.0	JM1- pre-B-cell lymphoma	5.7
Cerebellum	9.3	Jurkat- T cell leukemia	5.7
NCI-H292- Mucoepidermoid lung carcinoma	57.8	TF-1- Erythroleukemia	62.0
DMS-114- Small cell lung cancer	0.6	HUT 78- T-cell lymphoma	29.7
DMS-79- Small cell lung cancer	70.2	U937- Histiocytic lymphoma	86.5
NCI-H146- Small cell lung cancer	20.0	KU-812- Myelogenous leukemia	87.1
NCI-H526- Small cell lung cancer	35.6	769-P- Clear cell renal carcinoma	8.8
NCI-N417- Small cell lung cancer	3.7	Caki-2- Clear cell renal carcinoma	26.2
NCI-H82- Small cell lung cancer	6.6	SW 839- Clear cell renal carcinoma	70.7
NCI-H157- Squamous cell lung cancer (metastasis)	0.8	G401- Wilms' tumor	10.2
NCI-H1155- Large cell lung cancer	15.3	Hs766T- Pancreatic carcinoma (LN metastasis)	33.9
NCI-H1299- Large cell lung cancer	14.5	CAPAN-1- Pancreatic adenocarcinoma (liver metastasis)	15.7
NCI-H727- Lung carcinoid	25.0	SU86.86- Pancreatic carcinoma (liver metastasis)	100.0
NCI-UMC-11- Lung carcinoid	31.2	BxPC-3- Pancreatic adenocarcinoma	10.9
LX-1- Small cell lung cancer	30.6	HPAC- Pancreatic adenocarcinoma	5.8
Colo-205- Colon cancer	15.1	MIA PaCa-2- Pancreatic carcinoma	0.1
KM12- Colon cancer	24.7	CFPAC-1- Pancreatic ductal adenocarcinoma	37.6
KM20L2- Colon cancer	33.0	PANC-1- Pancreatic	2.9

		epithelioid ductal carcinoma	
NCI-H716- Colon cancer	24.1	T24- Bladder carcinoma (transitional cell)	12.4
SW-48- Colon adenocarcinoma	52.9	5637- Bladder carcinoma	9.0
SW1116- Colon adenocarcinoma	50.0	HT-1197- Bladder carcinoma	46.0
LS 174T- Colon adenocarcinoma	78.5	UM-UC-3- Bladder carcinoma (transitional cell)	5.5
SW-948- Colon adenocarcinoma	5.5	A204- Rhabdomyosarcoma	8.8
SW-480- Colon adenocarcinoma	25.9	HT-1080- Fibrosarcoma	10.4
NCI-SNU-5- Gastric carcinoma	15.2	MG-63- Osteosarcoma	6.7
KATO III- Gastric carcinoma	66.0	SK-LMS-1- Leiomyosarcoma (vulva)	13.2
NCI-SNU-16- Gastric carcinoma	20.6	SJRH30- Rhabdomyosarcoma (met to bone marrow)	4.7
NCI-SNU-1- Gastric carcinoma	85.3	A431- Epidermoid carcinoma	12.1
RF-1- Gastric adenocarcinoma	64.2	WM266-4- Melanoma	6.2
RF-48- Gastric adenocarcinoma	70.2	DU 145- Prostate carcinoma (brain metastasis)	0.0
MKN-45- Gastric carcinoma	33.9	MDA-MB-468- Breast adenocarcinoma	6.7
NCI-N87- Gastric carcinoma	28.5	SCC-4- Squamous cell carcinoma of tongue	0.9
OVCAR-5- Ovarian carcinoma	11.5	SCC-9- Squamous cell carcinoma of tongue	10.5
RL95-2- Uterine carcinoma	15.7	SCC-15- Squamous cell carcinoma of tongue	0.6
HelaS3- Cervical adenocarcinoma	10.5	CAL 27- Squamous cell carcinoma of tongue	27.4

Table AQF. Panel 4D

Tissue Name	Rel. Exp.(%) Ag3488, Run 166441742	Tissue Name	Rel. Exp.(%) Ag3488, Run 166441742
Secondary Th1 act	18.7	HUVEC IL-1beta	2.6
Secondary Th2 act	25.2	HUVEC IFN gamma	3.0
Secondary Tr1 act	29.5	HUVEC TNF alpha +	3.2

		IFN gamma	
Secondary Th1 rest	37.9	HUVEC TNF alpha + IL4	3.6
Secondary Th2 rest	21.3	HUVEC IL-11	3.5
Secondary Tr1 rest	29.3	Lung Microvascular EC none	9.2
Primary Th1 act	7.1	Lung Microvascular EC TNFalpha + IL-1beta	7.5
Primary Th2 act	20.4	Microvascular Dermal EC none	9.0
Primary Tr1 act	25.9	Microvascular Dermal EC TNFalpha + IL-1beta	4.4
Primary Th1 rest	95.9	Bronchial epithelium TNFalpha + IL1beta	6.5
Primary Th2 rest	55.1	Small airway epithelium none	6.0
Primary Tr1 rest	28.5	Small airway epithelium TNFalpha + IL-1beta	25.3
CD45RA CD4 lymphocyte act	8.8	Coronary artery SMC rest	7.5
CD45RO CD4 lymphocyte act	25.2	Coronary artery SMC TNFalpha + IL-1beta	4.0
CD8 lymphocyte act	12.6	Astrocytes rest	5.5
Secondary CD8 lymphocyte rest	31.2	Astrocytes TNFalpha + IL-1beta	12.2
Secondary CD8 lymphocyte act	7.6	KU-812 (Basophil) rest	41.5
CD4 lymphocyte none	50.3	KU-812 (Basophil) PMA/ionomycin	91.4
2ry Th1/Th2/Tr1_anti-CD95 CH11	41.8	CCD1106 (Keratinocytes) none	3.6
LAK cells rest	23.2	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	7.1
LAK cells IL-2	33.9	Liver cirrhosis	25.3
LAK cells IL-2+IL-12	26.4	Lupus kidney	21.6
LAK cells IL-2+IFN gamma	42.9	NCI-H292 none	46.0
LAK cells IL-2+ IL-18	24.0	NCI-H292 IL-4	43.8
LAK cells PMA/ionomycin	14.3	NCI-H292 IL-9	51.1
NK Cells IL-2 rest	14.2	NCI-H292 IL-13	26.2
Two Way MLR 3 day	39.8	NCI-H292 IFN gamma	23.5
Two Way MLR 5 day	18.7	HPAEC none	3.8
Two Way MLR 7 day	16.6	HPAEC TNF alpha + IL-	10.2

		1 beta	
PBMC rest	45.1	Lung fibroblast none	7.3
PBMC PWM	17.2	Lung fibroblast TNF alpha + IL-1 beta	11.7
PBMC PHA-L	19.8	Lung fibroblast IL-4	6.8
Ramos (B cell) none	23.8	Lung fibroblast IL-9	4.6
Ramos (B cell) ionomycin	18.0	Lung fibroblast IL-13	4.8
B lymphocytes PWM	21.8	Lung fibroblast IFN gamma	5.7
B lymphocytes CD40L and IL-4	43.2	Dermal fibroblast CCD1070 rest	8.7
EOL-1 dbcAMP	53.6	Dermal fibroblast CCD1070 TNF alpha	20.9
EOL-1 dbcAMP PMA/ionomycin	25.0	Dermal fibroblast CCD1070 IL-1 beta	3.3
Dendritic cells none	72.2	Dermal fibroblast IFN gamma	3.2
Dendritic cells LPS	29.1	Dermal fibroblast IL-4	7.0
Dendritic cells anti- CD40	80.7	IBD Colitis 2	17.6
Monocytes rest	100.0	IBD Crohn's	11.4
Monocytes LPS	11.0	Colon	93.3
Macrophages rest	92.0	Lung	27.4
Macrophages LPS	26.8	Thymus	17.6
HUVEC none	11.0	Kidney	56.6
HUVEC starved	9.7		

Table AQG. general oncology screening panel\_v\_2.4

Tissue Name	Rel. Exp.(%) Ag3488, Run 259737914	Tissue Name	Rel. Exp.(%) Ag3488, Run 259737914
Colon cancer 1.	6.9	Bladder cancer NAT 2	0.3
Colon cancer NAT 1	2.9	Bladder cancer NAT 3	0.2
Colon cancer 2	4.4	Bladder cancer NAT 4	0.8
Colon cancer NAT 2	2.6	Prostate adenocarcinoma 1	4.2
Colon cancer 3	27.4	Prostate adenocarcinoma 2	0.8
Colon cancer NAT 3	3.5	Prostate adenocarcinoma 3	1.8

Colon malignant cancer 4	12.6	Prostate adenocarcinoma 4	6.7
Colon normal adjacent tissue 4	1.1	Prostate cancer NAT 5	2.7
Lung cancer 1	2.7	Prostate adenocarcinoma 6	1.7
Lung NAT 1	0.5	Prostate adenocarcinoma 7	2.4
Lung cancer 2	11.8	Prostate adenocarcinoma 8	0.6
Lung NAT 2	0.6	Prostate adenocarcinoma 9	3.0
Squamous cell carcinoma 3	5.8	Prostate cancer NAT 10	0.3
Lung NAT 3	0.2	Kidney cancer 1	11.3
metastatic melanoma 1	3.2	Kidney NAT 1	1.1
Melanoma 2	0.8	Kidney cancer 2	55.1
Melanoma 3	0.7	Kidney NAT 2	2.8
metastatic melanoma 4	6.2	Kidney cancer 3	100.0
metastatic melanoma 5	4.7	Kidney NAT 3	0.6
Bladder cancer 1	0.7	Kidney cancer 4	31.6
Bladder cancer NAT 1	0.0	Kidney NAT 4	0.8
Bladder cancer 2	0.9		

**CNS\_neurodegeneration\_v1.0 Summary:** Ag3488 This panel does not show differential expression of this gene in Alzheimer's disease. However, this profile confirms the expression of this gene at moderate levels in the brain. Please see Panel 1.4 for discussion of utility of this gene in the central nervous system.

- 5 **General\_screening\_panel\_v1.4 Summary:** Ag3488 Highest expression of this gene is seen in a renal cancer cell line (CT=23.2). This gene is widely expressed in this panel, with high to moderate levels of expression seen in brain, colon, gastric, lung, breast, ovarian, and melanoma cancer cell lines. This expression profile suggests a role for this gene product in cell survival and proliferation. Modulation of this gene product may be useful in
- 10 the treatment of cancer.

Among tissues with metabolic function, this gene is expressed at high to moderate levels in pituitary, adipose, adrenal gland, pancreas, thyroid, and adult and fetal skeletal muscle, heart, and liver. This widespread expression among these tissues suggests that this gene product may play a role in normal neuroendocrine and metabolic function and that

5   disregulated expression of this gene may contribute to neuroendocrine disorders or metabolic diseases, such as obesity and diabetes.

This gene is also expressed at moderate levels in the CNS, including the hippocampus, thalamus, substantia nigra, amygdala, cerebellum and cerebral cortex. Therefore, therapeutic modulation of the expression or function of this gene may be useful in the

10   treatment of neurologic disorders, such as Alzheimer's disease, Parkinson's disease, schizophrenia, multiple sclerosis, stroke and epilepsy.

**Panel 2.2 Summary:** Ag3488 Highest expression is seen in a kidney cancer (CT=28). In addition, this gene is more highly expressed in kidney cancer than in the corresponding normal adjacent tissue. Thus, expression of this gene could be used as a marker of this

15   cancer. Furthermore, therapeutic modulation of the expression or function of this gene product may be useful in the treatment of kidney cancer.

**Panel 3D Summary:** Ag3488 Highest expression is seen in a pancreatic cancer cell line (CT=29.6). Moderate levels of expression are also seen in many cancer cell lines on this panel. Please see Panel 1.4 for discussion of utility of this gene in cancer.

20   **Panel 4D Summary:** Ag3488 Highest expression is seen in resting monocytes (CT=25.3). This gene is also expressed at moderate levels in a wide range of cell types of significance in the immune response in health and disease. These cells include members of the T-cell, B-cell, endothelial cell, macrophage/monocyte, and peripheral blood mononuclear cell family, as well as epithelial and fibroblast cell types from lung and skin, and normal tissues

25   represented by colon, lung, thymus and kidney. This ubiquitous pattern of expression suggests that this gene product may be involved in homeostatic processes for these and other cell types and tissues. This pattern is in agreement with the expression profile in General\_screening\_panel\_v1.4 and also suggests a role for the gene product in cell survival and proliferation. Therefore, modulation of the gene product with a functional therapeutic

30   may lead to the alteration of functions associated with these cell types and lead to improvement of the symptoms of patients suffering from autoimmune and inflammatory

diseases such as asthma, allergies, inflammatory bowel disease, lupus erythematosus, psoriasis, rheumatoid arthritis, and osteoarthritis.

- General oncology screening panel\_v\_2.4 Summary:** Ag3488 Highest expression is seen in kidney cancer (CT=23.2). In addition, this gene is more highly expressed in colon and
- 5 kidney cancer than in the corresponding normal adjacent tissue. Thus, expression of this gene could be used as a marker of these cancers. Furthermore, therapeutic modulation of the expression or function of this gene product may be useful in the treatment of colon and kidney cancer.

**AR. CG92142-01: GLYCEROL-3-PHOSPHATE ACYLTRANSFERASE.**

- 10 Expression of gene CG92142-01 was assessed using the primer-probe set Ag3774, described in Table ARA. Results of the RTQ-PCR runs are shown in Tables ARB, ARC, ARD, ARE and ARF.

Table ARA. Probe Name Ag3774

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-ggtgctgctaaaactgttcaac-3'	22	673	612
Probe	TET-5'-tggaacattcaaattcacaagggtca-3'-TAMRA	26	704	613
Reverse	5'-attcgtctcagttgcagcttt-3'	21	743	614

Table ARB. CNS\_neurodegeneration\_v1.0

Tissue Name	Rel. Exp.(%) Ag3774, Run 206871268	Tissue Name	Rel. Exp.(%) Ag3774, Run 206871268
AD 1 Hippo	29.1	Control (Path) 3 Temporal Ctx	29.3
AD 2 Hippo	73.7	Control (Path) 4 Temporal Ctx	50.3
AD 3 Hippo	10.0	AD 1 Occipital Ctx	22.4
AD 4 Hippo	14.6	AD 2 Occipital Ctx (Missing)	0.0
AD 5 Hippo	92.0	AD 3 Occipital Ctx	20.3
AD 6 Hippo	45.1	AD 4 Occipital Ctx	33.9
Control 2 Hippo	44.1	AD 5 Occipital Ctx	37.6
Control 4 Hippo	20.3	AD 6 Occipital Ctx	24.7
Control (Path) 3	19.9	Control 1 Occipital	11.3

Hippo		Ctx	
AD 1 Temporal Ctx	20.6	Control 2 Occipital Ctx	48.0
AD 2 Temporal Ctx	75.3	Control 3 Occipital Ctx	43.5
AD 3 Temporal Ctx	13.4	Control 4 Occipital Ctx	21.2
AD 4 Temporal Ctx	45.1	Control (Path) 1 Occipital Ctx	81.8
AD 5 Inf Temporal Ctx	100.0	Control (Path) 2 Occipital Ctx	12.9
AD 5 Sup Temporal Ctx	78.5	Control (Path) 3 Occipital Ctx	13.6
AD 6 Inf Temporal Ctx	43.5	Control (Path) 4 Occipital Ctx	45.1
AD 6 Sup Temporal Ctx	50.7	Control 1 Parietal Ctx	25.2
Control 1 Temporal Ctx	25.5	Control 2 Parietal Ctx	84.7
Control 2 Temporal Ctx	46.7	Control 3 Parietal Ctx	41.2
Control 3 Temporal Ctx	57.0	Control (Path) 1 Parietal Ctx	91.4
Control 3 Temporal Ctx	25.2	Control (Path) 2 Parietal Ctx	38.2
Control (Path) 1 Temporal Ctx	66.4	Control (Path) 3 Parietal Ctx	19.1
Control (Path) 2 Temporal Ctx	52.1	Control (Path) 4 Parietal Ctx	48.0

Table ARC. General\_screening\_panel\_v1.4

Tissue Name	Rel. Exp.(%) Ag3774, Run 213515543	Tissue Name	Rel. Exp.(%) Ag3774, Run 213515543
Adipose	63.7	Renal ca. TK-10	21.5
Melanoma* Hs688(A).T	16.0	Bladder	6.3
Melanoma* Hs688(B).T	74.7	Gastric ca. (liver met.) NCI-N87	9.7
Melanoma* M14	10.2	Gastric ca. KATO III	16.5
Melanoma* LOXIMVI	76.8	Colon ca. SW-948	3.3
Melanoma* SK- MEL-5	23.8	Colon ca. SW480	12.9

Squamous cell carcinoma SCC-4	5.8	Colon ca.* (SW480 met) SW620	8.6
Testis Pool	12.8	Colon ca. HT29	4.1
Prostate ca.* (bone met) PC-3	10.3	Colon ca. HCT-116	25.3
Prostate Pool	2.3	Colon ca. CaCo-2	52.5
Placenta	1.3	Colon cancer tissue	10.4
Uterus Pool	1.6	Colon ca. SW1116	3.0
Ovarian ca. OVCAR-3	10.6	Colon ca. Colo-205	2.9
Ovarian ca. SK-OV-3	15.6	Colon ca. SW-48	2.5
Ovarian ca. OVCAR-4	5.4	Colon Pool	4.5
Ovarian ca. OVCAR-5	6.3	Small Intestine Pool	5.9
Ovarian ca. IGROV-1	5.5	Stomach Pool	3.3
Ovarian ca. OVCAR-8	4.9	Bone Marrow Pool	2.8
Ovary	4.0	Fetal Heart	3.1
Breast ca. MCF-7	11.7	Heart Pool	4.0
Breast ca. MDA-MB-231	8.5	Lymph Node Pool	7.2
Breast ca. BT 549	6.5	Fetal Skeletal Muscle	11.0
Breast ca. T47D	8.9	Skeletal Muscle Pool	10.9
Breast ca. MDA-N	10.7	Spleen Pool	5.3
Breast Pool	5.0	Thymus Pool	7.6
Trachea	10.6	CNS cancer (glio/astro) U87-MG	9.7
Lung	1.0	CNS cancer (glio/astro) U-118-MG	19.1
Fetal Lung	6.2	CNS cancer (neuro;met) SK-N-AS	22.1
Lung ca. NCI-N417	3.2	CNS cancer (astro) SF-539	5.9
Lung ca. LX-1	9.3	CNS cancer (astro) SNB-75	22.5
Lung ca. NCI-H146	2.9	CNS cancer (glio) SNB-19	5.0
Lung ca. SHP-77	16.2	CNS cancer (glio) SF-295	100.0
Lung ca. A549	7.6	Brain (Amygdala) Pool	2.9
Lung ca. NCI-H526	1.9	Brain (cerebellum)	2.4

Lung ca. NCI-H23	12.7	Brain (fetal)	17.9
Lung ca. NCI-H460	7.7	Brain (Hippocampus) Pool	5.9
Lung ca. HOP-62	6.0	Cerebral Cortex Pool	7.5
Lung ca. NCI-H522	17.6	Brain (Substantia nigra) Pool	5.8
Liver	16.3	Brain (Thalamus) Pool	8.1
Fetal Liver	70.7	Brain (whole)	8.4
Liver ca. HepG2	42.9	Spinal Cord Pool	4.8
Kidney Pool	8.5	Adrenal Gland	65.5
Fetal Kidney	6.6	Pituitary gland Pool	1.0
Renal ca. 786-0	10.3	Salivary Gland	3.0
Renal ca. A498	2.5	Thyroid (female)	3.8
Renal ca. ACHN	7.3	Pancreatic ca. CAPAN2	5.4
Renal ca. UO-31	7.2	Pancreas Pool	5.7

Table ARD. Panel 2.2

Tissue Name	Rel. Exp.(%) Ag3774, Run 174448446	Tissue Name	Rel. Exp.(%) Ag3774, Run 174448446
Normal Colon	7.9	Kidney Margin (OD04348)	8.7
Colon cancer (OD06064)	4.9	Kidney malignant cancer (OD06204B)	2.2
Colon Margin (OD06064)	3.6	Kidney normal adjacent tissue (OD06204E)	0.4
Colon cancer (OD06159)	0.2	Kidney Cancer (OD04450-01)	3.4
Colon Margin (OD06159)	2.8	Kidney Margin (OD04450-03)	3.3
Colon cancer (OD06297-04)	0.6	Kidney Cancer 8120613	0.8
Colon Margin (OD06297-05)	2.3	Kidney Margin 8120614	1.0
CC Gr.2 ascend colon (ODO3921)	0.5	Kidney Cancer 9010320	1.6
CC Margin (ODO3921)	1.0	Kidney Margin 9010321	0.2
Colon cancer metastasis (OD06104)	1.6	Kidney Cancer 8120607	0.8
Lung Margin	1.1	Kidney Margin	0.3

(OD06104)		8120608	
Colon mets to lung (OD04451-01)	2.2	Normal Uterus	5.0
Lung Margin (OD04451-02)	2.3	Uterine Cancer 064011	1.1
Normal Prostate	0.6	Normal Thyroid	0.3
Prostate Cancer (OD04410)	1.2	Thyroid Cancer 064010	0.6
Prostate Margin (OD04410)	1.2	Thyroid Cancer A302152	2.2
Normal Ovary	1.0	Thyroid Margin A302153	2.9
Ovarian cancer (OD06283-03)	1.0	Normal Breast	61.6
Ovarian Margin (OD06283-07)	10.1	Breast Cancer (OD04566)	2.7
Ovarian Cancer 064008	3.3	Breast Cancer 1024	4.8
Ovarian cancer (OD06145)	2.1	Breast Cancer (OD04590-01)	4.8
Ovarian Margin (OD06145)	2.4	Breast Cancer Mets (OD04590-03)	30.1
Ovarian cancer (OD06455-03)	1.7	Breast Cancer Metastasis (OD04655- 05)	6.0
Ovarian Margin (OD06455-07)	1.3	Breast Cancer 064006	2.0
Normal Lung	3.1	Breast Cancer 9100266	1.5
Invasive poor diff. lung adeno (ODO4945-01)	1.4	Breast Margin 9100265	3.6
Lung Margin (ODO4945-03)	2.2	Breast Cancer A209073	1.1
Lung Malignant Cancer (OD03126)	2.0	Breast Margin A2090734	5.8
Lung Margin (OD03126)	0.7	Breast cancer (OD06083)	4.2
Lung Cancer (OD05014A)	1.2	Breast cancer node metastasis (OD06083)	12.6
Lung Margin (OD05014B)	7.1	Normal Liver	87.7
Lung cancer (OD06081)	0.1	Liver Cancer 1026	12.5
Lung Margin (OD06081)	2.0	Liver Cancer 1025	100.0
Lung Cancer (OD04237-01)	1.0	Liver Cancer 6004-T	63.7

Lung Margin (OD04237-02)	2.6	Liver Tissue 6004-N	4.8
Ocular Melanoma Metastasis	7.5	Liver Cancer 6005-T	28.5
Ocular Melanoma Margin (Liver)	19.5	Liver Tissue 6005-N	67.8
Melanoma Metastasis	2.0	Liver Cancer 064003	12.2
Melanoma Margin (Lung)	3.6	Normal Bladder	2.3
Normal Kidney	1.6	Bladder Cancer 1023	0.3
Kidney Ca, Nuclear grade 2 (OD04338)	3.3	Bladder Cancer A302173	1.4
Kidney Margin (OD04338)	1.3	Normal Stomach	6.0
Kidney Ca Nuclear grade 1/2 (OD04339)	2.2	Gastric Cancer 9060397	0.9
Kidney Margin (OD04339)	2.2	Stomach Margin 9060396	1.7
Kidney Ca, Clear cell type (OD04340)	0.7	Gastric Cancer 9060395	1.9
Kidney Margin (OD04340)	4.0	Stomach Margin 9060394	2.3
Kidney Ca, Nuclear grade 3 (OD04348)	0.9	Gastric Cancer 064005	1.9

Table ARE. Panel 4.1D

Tissue Name	Rel. Exp.(%) Ag3774, Run 170130276	Tissue Name	Rel. Exp.(%) Ag3774, Run 170130276
Secondary Th1 act	39.8	HUVEC IL-1beta	38.2
Secondary Th2 act	44.4	HUVEC IFN gamma	39.0
Secondary Tr1 act	33.7	HUVEC TNF alpha + IFN gamma	19.1
Secondary Th1 rest	9.5	HUVEC TNF alpha + IL4	28.1
Secondary Th2 rest	11.4	HUVEC IL-11	25.2
Secondary Tr1 rest	12.2	Lung Microvascular EC none	32.3
Primary Th1 act	36.6	Lung Microvascular EC TNFalpha + IL-1beta	36.3
Primary Th2 act	39.8	Microvascular Dermal EC none	26.4
Primary Tr1 act	28.9	Microvascular Dermal EC TNFalpha + IL-1beta	23.3

Primary Th1 rest	24.8	Bronchial epithelium TNFalpha + IL1beta	38.4
Primary Th2 rest	11.7	Small airway epithelium none	24.1
Primary Tr1 rest	23.2	Small airway epithelium TNFalpha + IL-1beta	28.9
CD45RA CD4 lymphocyte act	45.1	Coronary artery SMC rest	31.4
CD45RO CD4 lymphocyte act	45.1	Coronary artery SMC TNFalpha + IL-1beta	24.5
CD8 lymphocyte act	49.0	Astrocytes rest	46.3
Secondary CD8 lymphocyte rest	31.2	Astrocytes TNFalpha + IL-1beta	12.1
Secondary CD8 lymphocyte act	22.1	KU-812 (Basophil) rest	37.9
CD4 lymphocyte none	11.0	KU-812 (Basophil) PMA/ionomycin	49.3
2ry Th1/Th2/Tr1_anti- CD95 CH11	15.9	CCD1106 (Keratinocytes) none	56.3
LAK cells rest	18.7	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	34.6
LAK cells IL-2	31.4	Liver cirrhosis	38.4
LAK cells IL-2+IL-12	25.3	NCI-H292 none	25.2
LAK cells IL-2+IFN gamma	46.7	NCI-H292 IL-4	36.3
LAK cells IL-2+ IL-18	32.8	NCI-H292 IL-9	47.6
LAK cells PMA/ionomycin	3.9	NCI-H292 IL-13	37.1
NK Cells IL-2 rest	30.8	NCI-H292 IFN gamma	49.3
Two Way MLR 3 day	23.3	HPAEC none	27.7
Two Way MLR 5 day	37.6	HPAEC TNF alpha + IL- 1 beta	31.9
Two Way MLR 7 day	17.8	Lung fibroblast none	44.1
PBMC rest	4.1	Lung fibroblast TNF alpha + IL-1 beta	17.0
PBMC PWM	35.4	Lung fibroblast IL-4	34.9
PBMC PHA-L	20.9	Lung fibroblast IL-9	62.4
Ramos (B cell) none	76.8	Lung fibroblast IL-13	42.0
Ramos (B cell) ionomycin	68.8	Lung fibroblast IFN gamma	25.2
B lymphocytes PWM	41.2	Dermal fibroblast CCD1070 rest	<b>100.0</b>
B lymphocytes CD40L	28.9	Dermal fibroblast	66.4

and IL-4		CCD1070 TNF alpha	
EOL-1 dbcAMP	17.4	Dermal fibroblast CCD1070 IL-1 beta	38.2
EOL-1 dbcAMP PMA/ionomycin	20.9	Dermal fibroblast IFN gamma	17.0
Dendritic cells none	21.0	Dermal fibroblast IL-4	47.3
Dendritic cells LPS	5.7	Dermal Fibroblasts rest	29.5
Dendritic cells anti- CD40	22.5	Neutrophils TNFa+LPS	0.0
Monocytes rest	7.9	Neutrophils rest	2.3
Monocytes LPS	2.6	Colon	15.4
Macrophages rest	22.2	Lung	23.8
Macrophages LPS	4.5	Thymus	68.3
HUVEC none	29.7	Kidney	49.3
HUVEC starved	34.6		

Table ARF. Panel 5D

Tissue Name	Rel. Exp.(%) Ag3774, Run 223675472	Tissue Name	Rel. Exp.(%) Ag3774, Run 223675472
97457_Patient- 02go_adipose	17.7	94709_Donor 2 AM - A_adipose	19.6
97476_Patient- 07sk_skeletal muscle	3.6	94710_Donor 2 AM - B_adipose	9.3
97477_Patient- 07ut_uterus	2.3	94711_Donor 2 AM - C_adipose	7.5
97478_Patient- 07pl_placenta	2.2	94712_Donor 2 AD - A_adipose	56.6
97481_Patient- 08sk_skeletal muscle	6.4	94713_Donor 2 AD - B_adipose	72.2
97482_Patient- 08ut_uterus	1.6	94714_Donor 2 AD - C_adipose	70.2
97483_Patient- 08pl_placenta	0.8	94742_Donor 3 U - A_Mesenchymal Stem Cells	1.6
97486_Patient- 09sk_skeletal muscle	0.5	94743_Donor 3 U - B_Mesenchymal Stem Cells	1.8
97487_Patient- 09ut_uterus	2.1	94730_Donor 3 AM - A_adipose	13.1
97488_Patient- 09pl_placenta	0.8	94731_Donor 3 AM - B_adipose	8.5
97492_Patient- 10ut_uterus	1.6	94732_Donor 3 AM - C_adipose	8.7
97493_Patient-	1.4	94733_Donor 3 AD - A_adipose	100.0

10pl_placenta			
97495_Patient-11go_adipose	10.4	94734_Donor 3 AD - B_adipose	62.9
97496_Patient-11sk_skeletal muscle	2.8	94735_Donor 3 AD - C_adipose	53.2
97497_Patient-11ut_uterus	2.1	77138_Liver_HepG2untreated	56.6
97498_Patient-11pl_placenta	1.8	73556_Heart_Cardiac stromal cells (primary)	0.5
97500_Patient-12go_adipose	13.5	81735_Small Intestine	2.3
97501_Patient-12sk_skeletal muscle	6.0	72409_Kidney_Proximal Convoluted Tubule	1.0
97502_Patient-12ut_uterus	2.6	82685_Small intestine Duodenum	1.6
97503_Patient-12pl_placenta	0.4	90650_Adrenal_Adrenocortical adenoma	4.6
94721_Donor 2 U - A_Mesenchymal Stem Cells	3.5	72410_Kidney_HRCE	3.3
94722_Donor 2 U - B_Mesenchymal Stem Cells	3.7	72411_Kidney_HRE	2.7
94723_Donor 2 U - C_Mesenchymal Stem Cells	2.7	73139_Uterus_Uterine smooth muscle cells	1.2

**CNS\_neurodegeneration\_v1.0 Summary:** Ag3774 This panel confirms the expression of the CG92142-01 gene at low levels in the brains of an independent group of individuals.

However, no differential expression of this gene was detected between Alzheimer's diseased postmortem brains and those of non-demented controls in this experiment. Please

5 see Panel 1.4 for a discussion of the potential utility of this gene in treatment of central nervous system disorders.

**General\_screening\_panel\_v1.4 Summary:** Ag3774 Highest expression of the CG92142-01 gene is detected in CNS cancer (glio) SF-295 cell line (CT=26). High expression of this gene is also in number of cancer cell lines (pancreatic, CNS, colon, gastric, renal, lung, 10 breast, ovarian, squamous cell carcinoma, prostate and melanoma). Therefore, therapeutic modulation of the activity of this gene or its protein product, through the use of small molecule drugs might be beneficial in the treatment of these cancers.

Among tissues with metabolic or endocrine function, this gene is expressed at high to moderate levels in pancreas, adipose, adrenal gland, thyroid, pituitary gland, skeletal muscle, heart, liver and the gastrointestinal tract. Therefore, therapeutic modulation of the activity of this gene may prove useful in the treatment of endocrine/metabolically related diseases, such as obesity and diabetes.

The CG92142-01 gene codes for mitochondrial glycerol-3-phosphate acyltransferase (GPAT). GPAT is an adipocyte determination and differentiation factor 1 (ADD1) and sterol regulatory element-binding protein-1 (SREBP-1) regulated differentiation gene (Ref.1). It is up-regulated by insulin and high-carbohydrate diets (Ref.2). GPAT up-regulation increases triglyceride (TG) synthesis and fat deposition. Inhibition of GPAT activity could lead to decreased TG synthesis and fat deposition. Troglitazone, a thiazolidinedione compound used to treat non-insulin-dependent diabetes mellitus (NIDDM), was shown to decrease GPAT activity and adipogenesis in ZDF rat islets (ref.3). Therefore, therapeutic modulation of the activity of this gene may prove useful in the treatment of diabetes.

In addition, this gene is expressed at moderate levels in all regions of the central nervous system examined, including amygdala, hippocampus, substantia nigra, thalamus, cerebellum, cerebral cortex, and spinal cord. Therefore, this gene may play a role in central nervous system disorders such as Alzheimer's disease, Parkinson's disease, epilepsy, multiple sclerosis, schizophrenia and depression.

#### References.

1. Ericsson J, Jackson SM, Kim JB, Spiegelman BM, Edwards PA. (1997) Identification of glycerol-3-phosphate acyltransferase as an adipocyte determination and differentiation factor 1- and sterol regulatory element-binding protein-responsive gene. J Biol Chem 272(11):7298-305. PMID: 9054427.
2. Dircks LK, Sul HS. (1997) Mammalian mitochondrial glycerol-3-phosphate acyltransferase. Biochim Biophys Acta 1348(1-2):17-26. PMID: 9370312

3. Shimabukuro M, Zhou YT, Lee Y, Unger RH. (1998) Troglitazone lowers islet fat and restores beta cell function of Zucker diabetic fatty rats. J Biol Chem 273(6):3547-50 PMID: 9452481.

**Panel 2.2 Summary:** Ag3774 Highest expression of the CG92142-01 gene is detected in liver cancer 1025 sample (CT=28.7). In addition, low to moderate expression of this gene is seen in number of cancer and normal samples used in this panel. Please see Panel 1.4 for a discussion of the potential utility of this gene.

**Panel 4.1D Summary:** Ag3774 Highest expression of the CG92142-01 gene is detected in resting dermal fibroblast CCD1070 (CT=31). This gene is expressed at low to moderate levels in a wide range of cell types of significance in the immune response in health and disease. These cells include members of the T-cell, B-cell, endothelial cell, macrophage/monocyte, and peripheral blood mononuclear cell family, as well as epithelial and fibroblast cell types from lung and skin, and normal tissues represented by colon, lung, thymus and kidney. This ubiquitous pattern of expression suggests that this gene product may be involved in homeostatic processes for these and other cell types and tissues. This pattern is in agreement with the expression profile in General\_screening\_panel\_v1.4 and also suggests a role for the gene product in cell survival and proliferation. Therefore, modulation of the gene product with a functional therapeutic may lead to the alteration of functions associated with these cell types and lead to improvement of the symptoms of patients suffering from autoimmune and inflammatory diseases such as asthma, allergies, inflammatory bowel disease, lupus erythematosus, psoriasis, rheumatoid arthritis, and osteoarthritis.

Interestingly, expression of this gene is stimulated in PWM treated PBMC cells (CT=32.5) as compared to resting PBMC (35.6). Therefore, expression of this gene can be used to distinguish between resting and stimulated PBMC cells.

**Panel 5D Summary:** Ag3774 Highest expression of the CG92142-01 gene is detected in 94733\_Donor 3 AD-A\_adipose sample(CT=27.6). In addition, high to moderated expression of this gene is also seen in number of adipose, small intestine, uterus, skeletal muscle, placenta and mesenchymal stem cell samples. Please see Panel 1.4 for a discussion of the potential utility of this gene.

**AS. CG98102-03: Diamine AcetylTransferase.**

Expression of gene CG98102-03 was assessed using the primer-probe sets Ag4695, Ag4700, Ag4705 and Ag5877, described in Tables ASA, ASB, ASC and ASD. Results of the RTQ-PCR runs are shown in Tables ASE, ASF and ASG.

**5 Table ASA. Probe Name Ag4695.**

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5' -gccagcctgactgagaaga-3'	19	968	615
Probe	TET-5' -agacgaatgaggaaccacctcctcct-3' - TAMRA	26	929	616
Reverse	5' -caacaatgctgtgtccttcc-3'	20	658	617

**Table ASB. Probe Name Ag4700**

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5' -caatctcagatgcagtttggga-3'	21	174	618
Probe	TET-5' -tcagatctttctccttgaatatctttcga-3' - TAMRA	29	142	619
Reverse	5' -agatcacaccaccttgttgttt-3'	22	119	620

**Table ASC. Probe Name Ag4705**

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5' -ggctaaatatgaatacatggaag-3'	23	781	621
Probe	TET-5' -ttttggagagcacccttttaccac-3' - TAMRA	25	716	622
Reverse	5' -atgctgtgtccttccg-3'	16	663	623

**Table ASD. Probe Name Ag5877**

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5' -aagaggtgcttctgatctgtcc-3'	22	428	624
Probe	TET-5' -tgaagaggggttgagactgttcaagatcg-3' - TAMRA	29	397	625
Reverse	5' -catctacagcagcactcctcac-3'	22	341	626

**Table ASE. General\_screening\_panel\_v1.4**

Tissue	Rel.	Rel.	Rel.	Tissue Name	Rel.	Rel.	Rel.
--------	------	------	------	-------------	------	------	------

Name	Exp.(%) Ag4695, Run 219997539	Exp.(%) Ag4700, Run 222825527	Exp.(%) Ag4705, Run 213821747		Exp.(%) Ag4695, Run 219997539	Exp.(%) Ag4700, Run 222825527	Exp.(%) Ag4705, Run 213821747
Adipose	16.8	45.7	12.6	Renal ca. TK-10	9.4	14.4	11.3
Melanoma* Hs688(A).T	2.8	1.2	2.8	Bladder	100.0	67.4	100.0
Melanoma* Hs688(B).T	3.1	1.3	2.0	Gastric ca. (liver met.) NCI-N87	7.3	10.6	8.1
Melanoma* M14	25.5	13.7	18.4	Gastric ca. KATO III	90.8	22.8	55.9
Melanoma* LOXIMVI	1.0	0.6	1.8	Colon ca. SW-948	6.3	3.4	2.0
Melanoma* SK-MEL-5	11.9	19.5	14.2	Colon ca. SW480	26.4	20.9	28.7
Squamous cell carcinoma SCC-4	3.1	2.3	0.8	Colon ca.* (SW480 met) SW620	35.4	50.0	38.2
Testis Pool	5.6	3.1	4.5	Colon ca. HT29	3.0	4.4	3.8
Prostate ca.* (bone met) PC-3	16.7	8.4	17.3	Colon ca. HCT-116	21.5	27.9	31.0
Prostate Pool	4.9	5.5	2.2	Colon ca. CaCo-2	12.9	7.5	13.8
Placenta	20.0	6.9	0.1	Colon cancer tissue	36.3	54.0	45.4
Uterus Pool	1.0	11.6	0.3	Colon ca. SW1116	0.4	1.1	1.0
Ovarian ca. OVCAR-3	4.2	6.4	4.7	Colon ca. Colo-205	13.1	4.0	5.6
Ovarian ca. SK-OV-3	7.5	8.5	9.3	Colon ca. SW-48	6.7	2.3	3.9
Ovarian ca. OVCAR-4	1.7	1.2	1.5	Colon Pool	5.1	12.2	4.8
Ovarian ca. OVCAR-5	8.0	27.9	9.2	Small Intestine Pool	1.5	12.4	1.9
Ovarian ca. IGROV-1	32.5	83.5	40.9	Stomach Pool	24.3	31.6	17.6
Ovarian ca. OVCAR-8	9.1	20.7	4.1	Bone Marrow Pool	2.3	17.7	1.4
Ovary	5.1	5.6	4.9	Fetal Heart	1.8	2.1	2.2

Breast ca. MCF-7	1.6	3.1	2.0	Heart Pool	1.9	6.9	2.0
Breast ca. MDA-MB- 231	2.6	10.6	2.9	Lymph Node Pool	6.6	20.0	8.3
Breast ca. BT 549	25.5	9.0	22.2	Fetal Skeletal Muscle	0.7	1.5	0.7
Breast ca. T47D	16.6	71.2	19.2	Skeletal Muscle Pool	0.7	2.0	0.9
Breast ca. MDA-N	33.4	46.7	40.9	Spleen Pool	5.2	25.3	8.7
Breast Pool	10.4	19.3	7.5	Thymus Pool	8.7	37.4	11.1
Trachea	41.5	20.4	38.2	CNS cancer (glio/astro) U87-MG	14.9	17.7	12.6
Lung	0.9	24.1	0.9	CNS cancer (glio/astro) U- 118-MG	16.7	12.1	18.0
Fetal Lung	80.1	82.9	65.1	CNS cancer (neuro;met) SK-N-AS	0.4	1.2	1.0
Lung ca. NCI-N417	0.2	0.1	0.3	CNS cancer (astro) SF-539	0.6	1.3	0.9
Lung ca. LX-1	50.7	82.4	53.6	CNS cancer (astro) SNB-75	63.3	83.5	64.6
Lung ca. NCI-H146	0.6	0.3	0.8	CNS cancer (glio) SNB-19	27.5	54.7	37.9
Lung ca. SHP-77	0.8	1.8	1.2	CNS cancer (glio) SF-295	50.7	72.7	66.0
Lung ca. A549	27.2	28.1	23.7	Brain (Amygdala) Pool	2.9	4.9	3.5
Lung ca. NCI-H526	0.8	1.1	1.1	Brain (cerebellum)	1.1	1.0	1.2
Lung ca. NCI-H23	43.2	100.0	66.9	Brain (fetal)	6.0	4.2	6.0
Lung ca. NCI-H460	0.6	8.5	1.0	Brain (Hippocampus) Pool	7.8	6.7	5.7
Lung ca. HOP-62	3.6	23.8	5.1	Cerebral Cortex Pool	3.6	5.9	6.9
Lung ca. NCI-H522	2.9	6.4	3.5	Brain (Substantia nigra) Pool	5.1	6.0	7.9
Liver	3.5	0.8	1.4	Brain	5.7	6.5	8.6

				(Thalamus) Pool			
Fetal Liver	20.6	5.4	14.0	Brain (whole)	5.4	2.7	11.2
Liver ca. HepG2	11.6	19.6	16.7	Spinal Cord Pool	6.2	10.1	7.0
Kidney Pool	6.1	36.6	0.0	Adrenal Gland	12.8	5.1	14.7
Fetal Kidney	5.4	5.6	0.2	Pituitary gland Pool	2.4	2.0	4.0
Renal ca. 786-0	13.3	9.0	8.1	Salivary Gland	4.1	0.9	5.4
Renal ca. A498	4.9	2.4	5.8	Thyroid (female)	23.8	10.4	5.6
Renal ca. ACHN	1.7	2.2	1.9	Pancreatic ca. CAPAN2	8.0	10.3	9.7
Renal ca. UO-31	34.6	11.2	5.1	Pancreas Pool	11.8	21.6	17.0

Table ASF. General\_screening\_panel\_v1.5

Tissue Name	Rel. Exp.(%) Ag5877, Run 248204736	Tissue Name	Rel. Exp.(%) Ag5877, Run 248204736
Adipose	41.2	Renal ca. TK-10	15.7
Melanoma* Hs688(A).T	3.9	Bladder	100.0
Melanoma* Hs688(B).T	5.5	Gastric ca. (liver met.) NCI-N87	17.1
Melanoma* M14	40.3	Gastric ca. KATO III	58.2
Melanoma* LOXIMVI	1.8	Colon ca. SW-948	6.6
Melanoma* SK- MEL-5	20.6	Colon ca. SW480	30.8
Squamous cell carcinoma SCC-4	7.1	Colon ca.* (SW480 met) SW620	62.4
Testis Pool	7.5	Colon ca. HT29	4.3
Prostate ca.* (bone met) PC-3	16.4	Colon ca. HCT-116	34.9
Prostate Pool	17.0	Colon ca. CaCo-2	12.4
Placenta	38.2	Colon cancer tissue	59.0
Uterus Pool	7.4	Colon ca. SW1116	1.5
Ovarian ca. OVCAR-3	6.0	Colon ca. Colo-205	6.3
Ovarian ca. SK-	8.8	Colon ca. SW-48	4.2

OV-3			
Ovarian ca. OVCAR-4	3.2	Colon Pool	8.4
Ovarian ca. OVCAR-5	22.5	Small Intestine Pool	2.4
Ovarian ca. IGROV-1	67.8	Stomach Pool	22.1
Ovarian ca. OVCAR-8	22.1	Bone Marrow Pool	6.4
Ovary	10.7	Fetal Heart	3.4
Breast ca. MCF-7	3.3	Heart Pool	4.5
Breast ca. MDA-MB-231	9.0	Lymph Node Pool	12.7
Breast ca. BT 549	18.3	Fetal Skeletal Muscle	1.5
Breast ca. T47D	14.2	Skeletal Muscle Pool	2.7
Breast ca. MDA-N	33.0	Spleen Pool	20.6
Breast Pool	13.8	Thymus Pool	21.0
Trachea	38.2	CNS cancer (glio/astro) U87-MG	20.9
Lung	4.1	CNS cancer (glio/astro) U-118-MG	15.5
Fetal Lung	95.9	CNS cancer (neuro;met) SK-N-AS	1.5
Lung ca. NCI-N417	0.3	CNS cancer (astro) SF-539	0.9
Lung ca. LX-1	84.1	CNS cancer (astro) SNB-75	74.2
Lung ca. NCI-H146	0.5	CNS cancer (glio) SNB-19	80.7
Lung ca. SHP-77	1.9	CNS cancer (glio) SF-295	66.0
Lung ca. A549	43.8	Brain (Amygdala) Pool	4.9
Lung ca. NCI-H526	0.7	Brain (cerebellum)	3.4
Lung ca. NCI-H23	77.9	Brain (fetal)	6.4
Lung ca. NCI-H460	9.9	Brain (Hippocampus) Pool	8.3
Lung ca. HOP-62	5.8	Cerebral Cortex Pool	6.0
Lung ca. NCI-H522	8.6	Brain (Substantia nigra) Pool	5.4
Liver	3.3	Brain (Thalamus) Pool	7.5
Fetal Liver	17.0	Brain (whole)	5.8
Liver ca. HepG2	21.3	Spinal Cord Pool	9.2
Kidney Pool	15.3	Adrenal Gland	15.9

Fetal Kidney	8.5	Pituitary gland Pool	5.6
Renal ca. 786-0	8.1	Salivary Gland	4.3
Renal ca. A498	6.3	Thyroid (female)	28.1
Renal ca. ACHN	2.6	Pancreatic ca. CAPAN2	13.7
Renal ca. UO-31	32.1	Pancreas Pool	22.8

Table ASG. Panel 5D

Tissue Name	Rel. Exp.(%) Ag4695, Run 200923963	Rel. Exp.(%) Ag4695, Run 204244772	Rel. Exp.(%) Ag4700, Run 200923964	Rel. Exp.(%) Ag4700, Run 204244775	Rel. Exp.(%) Ag4705, Run 204245092
97457_Patient-02go_adipose	21.5	23.3	77.9	94.6	24.1
97476_Patient-07sk_skeletal muscle	3.5	4.5	52.1	47.3	4.9
97477_Patient-07ut_uterus	8.7	7.1	25.9	18.0	6.6
97478_Patient-07pl_placenta	66.9	69.7	100.0	100.0	69.7
97481_Patient-08sk_skeletal muscle	1.0	1.1	66.4	72.2	3.0
97482_Patient-08ut_uterus	1.6	8.0	10.9	7.2	7.4
97483_Patient-08pl_placenta	30.1	30.6	39.2	54.0	26.6
97486_Patient-09sk_skeletal muscle	0.8	0.5	9.7	10.2	0.5
97487_Patient-09ut_uterus	4.9	3.1	21.2	14.5	4.3
97488_Patient-09pl_placenta	35.6	54.7	77.9	65.1	47.3
97492_Patient-10ut_uterus	8.8	10.7	34.2	25.5	8.3
97493_Patient-10pl_placenta	100.0	100.0	79.0	97.9	100.0
97495_Patient-11go_adipose	7.2	7.0	40.9	36.3	6.9
97496_Patient-11sk_skeletal muscle	0.9	0.8	12.3	6.7	1.7
97497_Patient-11ut_uterus	10.8	10.2	17.1	27.0	23.7
97498_Patient-11pl_placenta	61.1	76.8	80.7	58.2	50.3
97500_Patient-12go_adipose	10.2	0.0	70.2	57.8	12.7
97501_Patient-12sk_skeletal muscle	1.8	1.7	17.9	21.6	2.8
97502_Patient-12ut_uterus	14.5	13.2	35.8	51.1	18.4
97503_Patient-12pl_placenta	72.2	70.7	72.7	52.5	68.8
94721_Donor 2 U - A_Mesenchymal Stem Cells	3.0	2.7	4.1	3.6	9.5
94722_Donor 2 U - B_Mesenchymal Stem Cells	2.1	2.9	3.6	3.3	3.3

94723_Donor 2 U - C_Mesenchymal Stem Cells	2.0	0.1	4.0	2.7	2.3
94709_Donor 2 AM - A_adipose	9.0	10.4	6.8	8.8	8.8
94710_Donor 2 AM - B_adipose	6.5	5.5	5.8	2.9	5.2
94711_Donor 2 AM - C_adipose	4.2	2.9	4.3	6.0	3.4
94712_Donor 2 AD - A_adipose	7.2	8.0	16.2	11.7	7.6
94713_Donor 2 AD - B_adipose	9.6	12.2	13.7	11.8	12.2
94714_Donor 2 AD - C_adipose	8.8	9.7	9.3	7.0	12.9
94742_Donor 3 U - A_Mesenchymal Stem Cells	1.0	0.7	2.2	1.2	1.1
94743_Donor 3 U - B_Mesenchymal Stem Cells	1.5	1.3	2.9	4.0	1.9
94730_Donor 3 AM - A_adipose	14.0	12.8	22.7	15.6	9.8
94731_Donor 3 AM - B_adipose	7.2	29.1	7.0	10.8	6.8
94732_Donor 3 AM - C_adipose	5.7	9.2	9.5	11.9	9.0
94733_Donor 3 AD - A_adipose	17.2	20.3	17.0	20.6	15.3
94734_Donor 3 AD - B_adipose	9.7	6.9	11.7	6.7	7.1
94735_Donor 3 AD - C_adipose	11.1	11.9	19.2	13.8	10.3
77138_Liver_HepG2untreated	27.5	27.5	34.2	39.2	23.3
73556_Heart_Cardiac stromal cells (primary)	3.5	3.0	10.0	8.0	7.2
81735_Small Intestine	13.3	12.1	49.0	48.0	15.5
72409_Kidney_Proximal Convolutd Tubule	5.8	5.1	15.0	8.4	5.6
82685_Small intestine Duodenum	17.9	19.5	60.3	44.8	28.1
90650_Adrenal_Adrenocortical adenoma	2.7	0.0	25.3	24.3	4.9
72410_Kidney_HRCE	30.1	33.4	39.0	38.7	25.0
72411_Kidney_HRE	28.5	23.2	40.9	50.0	22.4
73139_Uterus_Uterine smooth muscle cells	2.0	1.1	4.5	3.9	1.4

- General\_screening\_panel\_v1.4 Summary:** Ag4695/Ag4700/Ag4705 Three experiments using three probe-primer sets gave results that are in good agreement. This gene is expressed at moderate to high levels in all of the tissues on this panel, with highest expression in bladder and a lung cancer cell line (CTs=24-28). Interestingly, expression of this gene is higher in fetal lung and lung cancer cell lines when compared to adult lung. Expression of this gene is also upregulated in colon cancer cell lines when compared to normal colon. Therefore, therapeutic modulation of the activity of this gene or its protein product, through the use of small molecule drugs, protein therapeutics or antibodies, might be beneficial in the treatment of lung and colon cancer.
- 10 In addition, this gene is expressed at moderate levels in all regions of the central nervous system examined, including amygdala, hippocampus, substantia nigra, thalamus, cerebellum, cerebral cortex, and spinal cord. Therefore, this gene may play a role in central nervous system disorders such as Alzheimer's disease, Parkinson's disease, epilepsy, multiple sclerosis, schizophrenia and depression.
- 15 Among tissues with metabolic or endocrine function, this gene is expressed at high to moderate levels in pancreas, adipose, adrenal gland, thyroid, pituitary gland, skeletal muscle, heart, liver and the gastrointestinal tract. Therefore, therapeutic modulation of the activity of this gene may prove useful in the treatment of endocrine/metabolically related diseases, such as obesity and diabetes.
- 20 **General\_screening\_panel\_v1.5 Summary:** Ag5877. Expression of this gene is highest in bladder (CT = 23.6). This gene is expressed at moderate to high levels in all of the tissues on this panel, consistent with what is observed in Panel 1.4. Interestingly, expression of this gene is higher in fetal lung (CT = 23.7) and a subset of lung cancer cell lines (CTs = 24) when compared to adult lung (CT = 28.2). Expression of this gene is also upregulated in colon cancer cell lines (CTs = 24) when compared to normal colon (CT = 27.2). Therefore, therapeutic modulation of the activity of this gene or its protein product, through the use of small molecule drugs, protein therapeutics or antibodies, might be beneficial in the treatment of lung and colon cancer. Please see Panel 1.4 for additional discussion of the potential relevance of this gene in human disease.
- 25
- 30 **Panel 5D Summary:** Ag4695/Ag4705 Three experiments using two probe-primer sets gave results that are in good agreement. This gene is expressed at moderate to high levels

in the majority of metabolic tissues on this panel, with highest expression in a placenta sample from a diabetic patient (CTs = 23-28). Ag4700 Two experiment with same probe-primer sets are in excellent agreement. This gene shows widespread expression with highest expression of this gene in placenta of non-diabetic patient (CTs=30-30.7).

- 5 Spermine has been demonstrated to enhance insulin receptor binding in a dose dependent manner [Pedersen et al., Mol Cell Endocrinol., 1989 Apr;62(2):161-6]. Thus, it was proposed that polyamines may act as intracellular or intercellular (autocrine) regulators to modulate insulin binding. It has also been shown that the insulin-like effects elicited by polyamines in fat cells (e.g. enhancement of glucose transport and inhibition of cAMP-
- 10 mediated lipolysis) are dependent on H<sub>2</sub>O<sub>2</sub> production (Livingston et al., J. Biol. Chem., 1977 Jan 25;252(2):560-2). Inhibiting polyamine catabolism through an inhibitor of this rate-limiting enzyme may abolish the insulin-like antilipolytic effects of polyamines. Therefore, therapeutic inhibition of the activity of this gene using small molecule drugs may be of benefit in the treatment of obesity.

15 **Example D: Identification of Single Nucleotide Polymorphisms in NOVX nucleic acid sequences**

- Variant sequences are also included in this application. A variant sequence can include a single nucleotide polymorphism (SNP). A SNP can, in some instances, be referred to as a "cSNP" to denote that the nucleotide sequence containing the SNP
- 20 originates as a cDNA. A SNP can arise in several ways. For example, a SNP may be due to a substitution of one nucleotide for another at the polymorphic site. Such a substitution can be either a transition or a transversion. A SNP can also arise from a deletion of a nucleotide or an insertion of a nucleotide, relative to a reference allele. In this case, the polymorphic site is a site at which one allele bears a gap with respect to a particular
- 25 nucleotide in another allele. SNPs occurring within genes may result in an alteration of the amino acid encoded by the gene at the position of the SNP. Intragenic SNPs may also be silent, when a codon including a SNP encodes the same amino acid as a result of the redundancy of the genetic code. SNPs occurring outside the region of a gene, or in an intron within a gene, do not result in changes in any amino acid sequence of a protein but
- 30 may result in altered regulation of the expression pattern. Examples include alteration in temporal expression, physiological response regulation, cell type expression regulation, intensity of expression, and stability of transcribed message.

SeqCalling assemblies produced by the exon linking process were selected and extended using the following criteria. Genomic clones having regions with 98% identity to all or part of the initial or extended sequence were identified by BLASTN searches using the relevant sequence to query human genomic databases. The genomic clones that  
5 resulted were selected for further analysis because this identity indicates that these clones contain the genomic locus for these SeqCalling assemblies. These sequences were analyzed for putative coding regions as well as for similarity to the known DNA and protein sequences. Programs used for these analyses include Grail, Genscan, BLAST, HMMER, FASTA, Hybrid and other relevant programs.

10 Some additional genomic regions may have also been identified because selected SeqCalling assemblies map to those regions. Such SeqCalling sequences may have overlapped with regions defined by homology or exon prediction. They may also be included because the location of the fragment was in the vicinity of genomic regions identified by similarity or exon prediction that had been included in the original predicted  
15 sequence. The sequence so identified was manually assembled and then may have been extended using one or more additional sequences taken from CuraGen Corporation's human SeqCalling database. SeqCalling fragments suitable for inclusion were identified by the CuraTools™ program SeqExtend or by identifying SeqCalling fragments mapping to the appropriate regions of the genomic clones analyzed.

20 The regions defined by the procedures described above were then manually integrated and corrected for apparent inconsistencies that may have arisen, for example, from miscalled bases in the original fragments or from discrepancies between predicted exon junctions, EST locations and regions of sequence similarity, to derive the final sequence disclosed herein. When necessary, the process to identify and analyze SeqCalling  
25 assemblies and genomic clones was reiterated to derive the full length sequence (Alderborn et al., Determination of Single Nucleotide Polymorphisms by Real-time Pyrophosphate DNA Sequencing. Genome Research. 10 (8) 1249-1265, 2000).

Variants are reported individually but any combination of all or a select subset of variants are also included as contemplated NOVX embodiments of the invention.

30. RESULTS:

**NOV 3b SNP Data**

Two polymorphic variants of NOV3b have been identified and are shown in Table 3S.

Table 3S						
Variant No.	Nucleotides			Amino Acids		
	Base Position of SNP	Wild-type	Variant	Base Position of SNP	Wild-type	Variant
13381488	314	C	T	65	Ser	Ser
13381501	803	G	T	228	Val	Val

5

**NOV 5b SNP Data**

One polymorphic variant of NOV5b has been identified and are shown in Table 5S.

Table 5S						
Variant No.	Nucleotides			Amino Acids		
	Base Position of SNP	Wild-type	Variant	Base Position of SNP	Wild-type	Variant
13381503	3017	G	A	999	Lys	Lys

10

**NOV 8a SNP Data**

Four polymorphic variants of NOV8a have been identified and are shown in Table 8S.

Table 8S						
Variant No.	Nucleotides			Amino Acids		
	Base Position of SNP	Wild-type	Variant	Base Position of SNP	Wild-type	Variant
c34c-cip1.113	981	G	C	324	Leu	Leu
13381270	1033	A	G	342	Met	Val
13381350	1042	A	G	345	Ile	Val
13376329	1222	T	C	405	Ser	Pro

15

**NOV 9a SNP Data**

Four polymorphic variants of NOV9a have been identified and are shown in

Table 9S.

Table 9S						
Variant No.	Nucleotides			Amino Acids		
	Base Position of SNP	Wild-type	Variant	Base Position of SNP	Wild-type	Variant
13381343	276	C	T	92	Phe	Phe
13381344	1045	G	T	349	Ala	Ser
13381348	1416	C	T	472	Gly	Gly
13381345	1802	G	C	601	Gly	Ala

5

**NOV 10a SNP Data**

One polymorphic variant of NOV10a has been identified and are shown in

Table 10S.

Table 10S						
Variant No.	Nucleotides			Amino Acids		
	Base Position of SNP	Wild-type	Variant	Base Position of SNP	Wild-type	Variant
13379513	1447	C	T	423	Pro	Pro

10

**NOV 12a SNP Data**

Two polymorphic variants of NOV12a have been identified and are shown in

15 Table 12S.

Table 12S						
Variant No.	Nucleotides			Amino Acids		
	Base Position of SNP	Wild-type	Variant	Base Position of SNP	Wild-type	Variant
13379505	139	C	T	15	Pro	Ser
13379506	221	C	T	42	Ser	Phe

**NOV 13a SNP Data**

Thirteen polymorphic variants of NOV13a have been identified and are shown in Table 13S.

Table 13S						
Variant No.	Nucleotides			Amino Acids		
	Base Position of SNP	Wild-type	Variant	Base Position of SNP	Wild-type	Variant
13376183	75	A	G	2	Gln	Gln
13376184	182	C	T	38	Ala	Val
13376185	184	G	A	39	Ala	Thr
13376186	223	A	G	52	Thr	Ala
13376187	256	C	T	63	Arg	Cys
13376188	328	A	G	87	Asn	Asp
13376189	347	C	T	93	Ala	Val
13376190	373	A	G	102	Thr	Ala
13376191	1257	C	T	396	Thr	Thr
13376192	1342	A	G	425	Ser	Gly
13376193	1549	G	A	494	Val	Met
13376194	1581	G	A	504	Thr	Thr
13381349	1607	A	G	513	Gln	Arg

5

**NOV 14a SNP Data**

One polymorphic variant of NOV14a has been identified and are shown in Table 14S.

Table 14S						
Variant No.	Nucleotides			Amino Acids		
	Base Position of SNP	Wild-type	Variant	Base Position of SNP	Wild-type	Variant
13376195	402	T	C	134	Ala	Ala

10

**NOV 19 SNP Data**

One polymorphic variant of NOV19 has been identified and are shown in Table 19S.

Table 19S						
Variant No.	Nucleotides			Amino Acids		
	Base Position of SNP	Wild-type	Variant	Base Position of SNP	Wild-type	Variant
13381369	1380	G	C	460	Ala	Ala

**NOV 20c SNP Data**

- 5 One polymorphic variant of NOV20c has been identified and are shown in

Table 20S.

Table 20S						
Variant No.	Nucleotides			Amino Acids		
	Base Position of SNP	Wild-type	Variant	Base Position of SNP	Wild-type	Variant
13381370	281	C	T	94	Thr	Met

**NOV 48a SNP Data**

- 10 One polymorphic variant of NOV48a has been identified and are shown in

Table 48S.

Table 48S						
Variant No.	Nucleotides			Amino Acids		
	Base Position of SNP	Wild-type	Variant	Base Position of SNP	Wild-type	Variant
13381473	532	C	G	145	Gln	Glu

**NOV 50a SNP Data**

- 15 Two polymorphic variants of NOV50a have been identified and are shown in

Table 50S.

Table 50S						
Variant No.	Nucleotides			Amino Acids		
	Base Position of SNP	Wild-type	Variant	Base Position of SNP	Wild-type	Variant
13381514	744	A	G	242	Ser	Gly

13381513	1009	T	C	330	Leu	Ser
----------	------	---	---	-----	-----	-----

**NOV 53b SNP Data**

Six polymorphic variants of NOV53b have been identified and are shown in

5 Table 53S.

Table 53S						
Variant No.	Nucleotides			Amino Acids		
	Base Position of SNP	Wild-type	Variant	Base Position of SNP	Wild-type	Variant
13374617	437	A	G	143	Asn	Ser
13375310	664	T	G	219	Phe	Val
13375309	1150	G	T	381	Ala	Ser
13375308	1210	G	T	401	Glu	End
13375307	1770	C	T	587	Asn	Asn
13374615	2011	A	G	0		

**NOV 54b SNP Data**

Two polymorphic variants of NOV54b have been identified and are shown in

10 Table 54S.

Table 54S						
Variant No.	Nucleotides			Amino Acids		
	Base Position of SNP	Wild-type	Variant	Base Position of SNP	Wild-type	Variant
13381471	472	G	A	145	Pro	Pro
13381470	1082	A	G	0		

**NOV 55a SNP Data**

One polymorphic variant of NOV55a has been identified and are shown in

15 Table 55S.

Table 55S						
Variant No.	Nucleotides			Amino Acids		
	Base Position of SNP	Wild-type	Variant	Base Position of SNP	Wild-type	Variant

13375795	1070	C	T	236	Arg	Trp
----------	------	---	---	-----	-----	-----

**NOV 56a SNP Data**

Six polymorphic variant of NOV56a has been identified and are shown in

5 Table 56S.

Table 56S						
Variant No.	Nucleotides			Amino Acids		
	Base Position of SNP	Wild-type	Variant	Base Position of SNP	Wild-type	Variant
13375586	430	T	C	110	Ser	Ser
13375585	492	A	G	131	Glu	Gly
13375583	1756	C	T	552	Asn	Asn
13375582	2143	T	A	681	Pro	Pro
13377559	2550	A	G	817	Lys	Arg
13377776	2555	C	T	819	Leu	Leu

**NOV 57b SNP Data**

Two polymorphic variants of NOV57b have been identified and are shown in

10 Table 57S.

Table 57S						
Variant No.	Nucleotides			Amino Acids		
	Base Position of SNP	Wild-type	Variant	Base Position of SNP	Wild-type	Variant
13376786	1433	G	A	455	Cys	Tyr
13376785	1435	A	G	456	Lys	Glu

**NOV 58a SNP Data**

Two polymorphic variant of NOV58a has been identified and are shown in

15 Table 58S.

Table 58S		
Variant	Nucleotides	Amino Acids

No.	Base Position of SNP	Wild-type	Variant	Base Position of SNP	Wild-type	Variant
13381335	499	G	A	145	Glu	Glu
13381336	1045	C	T	327	Asn	Asn

**NOV 59b SNP Data**

Three polymorphic variant of NOV59b has been identified and are shown in

5 Table 59S.

Table 59S						
Variant No.	Nucleotides			Amino Acids		
	Base Position of SNP	Wild-type	Variant	Base Position of SNP	Wild-type	Variant
13379479	21	T	C	0		
13381483	183	C	T	2	Ala	Val
13381482	520	C	T	114	Ser	Ser

**Example E. Method of Use**

The present invention is partially based on the identification of biological macromolecules differentially modulated in a pathologic state, disease, or an abnormal condition or state, and/or based on novel associations of proteins and polypeptides and the nucleic acids that encode them, as identified in a yeast 2-hybrid screen using a cDNA library or one-by-one matrix reactions. Among the pathologies or diseases of present interest include metabolic diseases including those related to endocrinologic disorders, cancers, various tumors and neoplasias, inflammatory disorders, central nervous system disorders, and similar abnormal conditions or states. Important metabolic disorders with which the biological macromolecules are associated include obesity and diabetes mellitus, especially obesity and Type II diabetes. It is believed that obesity predisposes a subject to Type II diabetes. In very significant embodiments of the present invention, the biological macromolecules implicated in these pathologies and conditions are proteins and polypeptides, and in such cases the present invention is related as well to the nucleic acids that encode them. Methods that may be employed to identify relevant biological macromolecules include any procedures that detect differential expression of nucleic acids encoding proteins and polypeptides associated with the disorder, as well as procedures that detect the respective proteins and polypeptides themselves. Significant methods that have been employed by the present inventors, include GeneCalling® technology and SeqCalling™ technology, disclosed respectively, in U. S. Patent No. 5,871,697, and in U. S. Ser. No. 09/417,386, filed Oct. 13, 1999, each of which is incorporated herein by reference in its entirety. GeneCalling® is also described in Shimkets, *et al.*, *Nature Biotechnology* 17:198-803 (1999).

The invention provides polypeptides and nucleotides encoded thereby that have been identified as having novel associations with a disease or pathology, or an abnormal state or condition, in a mammal. Included in the invention are nucleic acid sequences and their encoded polypeptides. The sequences are collectively referred to as “obesity and/or diabetes nucleic acids” or “obesity and/or diabetes polynucleotides” and the corresponding encoded polypeptide is referred to as an “obesity and/or diabetes polypeptide” or “obesity and/or diabetes protein”. For example, an obesity and/or diabetes nucleic acid according to the invention is a nucleic acid including an obesity and/or diabetes nucleic acid, and an obesity and/or diabetes polypeptide according to the invention is a polypeptide that includes the amino acid sequence of an obesity and/or diabetes polypeptide. Unless

indicated otherwise, "obesity and/or diabetes" is meant to refer to any of the sequences having novel associations disclosed herein.

The present invention identifies a set of proteins and polypeptides, including naturally occurring polypeptides, precursor forms or proproteins, or mature forms of the polypeptides or proteins, which are implicated as targets for therapeutic agents in the treatment of various diseases, pathologies, abnormal states and conditions. A target may be employed in any of a variety of screening methodologies in order to identify candidate therapeutic agents which interact with the target and in so doing exert a desired or favorable effect. The candidate therapeutic agent is identified by screening a large collection of substances or compounds in an important embodiment of the invention. Such a collection may comprise a combinatorial library of substances or compounds in which, in at least one subset of substances or compounds, the individual members are related to each other by simple structural variations based on a particular canonical or basic chemical structure. The variations may include, by way of nonlimiting example, changes in length or identity of a basic framework of bonded atoms; changes in number, composition and disposition of ringed structures, bridge structures, alicyclic rings, and aromatic rings; and changes in pendent or substituents atoms or groups that are bonded at particular positions to the basic framework of bonded atoms or to the ringed structures, the bridge structures, the alicyclic structures, or the aromatic structures.

The present invention discloses novel associations of proteins and polypeptides and the nucleic acids that encode them, as identified in a yeast 2-hybrid screen using a cDNA library or one-by-one matrix reactions. The proteins and related proteins that are similar to them are encoded by a cDNA and/or by genomic DNA and were identified in some cases by CuraGen Corporation.

In the current invention, protein interactions may include the interaction of a protein fragment with full-length protein, a protein fragment with another protein fragment, or full-length proteins with each other. The protein interactions disclosed in the present invention may also represent significant discoveries of functional importance to specific diseases or pathological conditions in which novel proteins are found to be components of known pathways, known proteins are found to be components of novel pathways, or novel proteins are found to be components of novel pathways.

A polypeptide or protein described herein, and that serves as a target in the screening procedure, includes the product of a naturally occurring polypeptide or precursor

form or proprotein. The naturally occurring polypeptide, precursor or proprotein includes, e.g., the full-length gene product, encoded by the corresponding gene. The naturally occurring polypeptide also includes the polypeptide, precursor or proprotein encoded by an open reading frame described herein. A “mature” form of a polypeptide or protein arises as  
5 a result of one or more naturally occurring processing steps as they may occur within the cell, including a host cell. The processing steps occur as the gene product arises, e.g., via cleavage of the amino-terminal methionine residue encoded by the initiation codon of an open reading frame, or the proteolytic cleavage of a signal peptide or leader sequence. Thus, a mature form arising from a precursor polypeptide or protein that has residues 1 to  
10 N, where residue 1 is the N-terminal methionine, would have residues 2 through N remaining. Alternatively, a mature form arising from a precursor polypeptide or protein having residues 1 to N, in which an amino-terminal signal sequence from residue 1 to residue M is cleaved, includes the residues from residue M+1 to residue N remaining. A “mature” form of a polypeptide or protein may also arise from non-proteolytic post-  
15 translational modification. Such non-proteolytic processes include, e.g., glycosylation, myristylation or phosphorylation. In general, a mature polypeptide or protein may result from the operation of only one of these processes, or the combination of any of them.

As used herein, “identical” residues correspond to those residues in a comparison between two sequences where the equivalent nucleotide base or amino acid residue in an  
20 alignment of two sequences is the same residue. Residues are alternatively described as “similar” or “positive” when the comparisons between two sequences in an alignment show that residues in an equivalent position in a comparison are either the same amino acid or a conserved amino acid as defined below.

As used herein, a “chemical composition” relates to a composition including at least  
25 one compound that is either synthesized or extracted from a natural source. A chemical compound may be the product of a defined synthetic procedure. Such a synthesized compound is understood herein to have defined properties in terms of molecular formula, molecular structure relating the association of bonded atoms to each other, physical properties such as electropherographic or spectroscopic characterizations, and the like. A  
30 compound extracted from a natural source is advantageously analyzed by chemical and physical methods in order to provide a representation of its defined properties, including its molecular formula, molecular structure relating the association of bonded atoms to each

other, physical properties such as electropherographic or spectroscopic characterizations, and the like.

As used herein, a “candidate therapeutic agent” is a chemical compound that includes at least one substance shown to bind to a target biopolymer. In important  
5       embodiments of the invention, the target biopolymer is a protein or polypeptide, a nucleic acid, a polysaccharide or proteoglycan, or a lipid such as a complex lipid. The method of identifying compounds that bind to the target effectively eliminates compounds with little or no binding affinity, thereby increasing the potential that the identified chemical  
10       compound may have beneficial therapeutic applications. In cases where the “candidate therapeutic agent” is a mixture of more than one chemical compound, subsequent screening procedures may be carried out to identify the particular substance in the mixture that is the binding compound, and that is to be identified as a candidate therapeutic agent.

As used herein, a “pharmaceutical agent” is provided by screening a candidate therapeutic agent using models for a disease state or pathology in order to identify a  
15       candidate exerting a desired or beneficial therapeutic effect with relation to the disease or pathology. Such a candidate that successfully provides such an effect is termed a pharmaceutical agent herein. Nonlimiting examples of model systems that may be used in such screens include particular cell lines, cultured cells, tissue preparations, whole tissues, organ preparations, intact organs, and nonhuman mammals. Screens employing at least  
20       one system, and preferably more than one system, may be employed in order to identify a pharmaceutical agent. Any pharmaceutical agent so identified may be pursued in further investigation using human subjects.

The following sections describe the study design(s) and the techniques used to identify these proteins, and any variants thereof, and to demonstrate its suitability as  
25       diagnostic markers, targets for an antibody therapeutic and targets for a small molecule drugs for Obesity and Diabetes.

## Methods

### 1. RTQ-PCR (Real Time Quantitative Polymerase Chain Reaction) Technology:

30       The quantitative expression of various clones was assessed using microtiter plates containing RNA samples from a variety of normal and pathology-derived cells, cell lines and tissues using real time quantitative PCR (RTQ PCR). RTQ PCR was performed on a Perkin-Elmer Biosystems ABI PRISM® 7700 Sequence Detection System. Various

collections of samples are assembled on the plates, and referred to as Panel 1 (containing cells and cell lines from normal and cancer sources), Panel 2 (containing samples derived from tissues, in particular from surgical samples, from normal and cancer sources), Panel 3 (containing samples derived from a wide variety of cancer sources), Panel 4 (containing  
5 cells and cell lines from normal cells and cells related to inflammatory conditions) and Panel CNSD.01 (containing samples from normal and diseased brains).

First, the RNA samples were normalized to reference nucleic acids such as constitutively expressed genes (for example,  $\beta$ -actin and GAPDH). Normalized RNA (5 ul) was converted to cDNA and analyzed by RTQ-PCR using One Step RT-PCR Master  
10 Mix Reagents (PE Biosystems; Catalog No. 4309169) and gene-specific primers according to the manufacturer's instructions. Probes and primers were designed for each assay according to Perkin Elmer Biosystem's *Primer Express* Software package (version I for Apple Computer's Macintosh Power PC) or a similar algorithm using the target sequence as input. Default settings were used for reaction conditions and the following parameters  
15 were set before selecting primers: primer concentration = 250 nM, primer melting temperature ( $T_m$ ) range = 58°-60° C, primer optimal  $T_m$  = 59° C, maximum primer difference = 2° C, probe does not have 5' G, probe  $T_m$  must be 10° C greater than primer  $T_m$ , amplicon size 75 bp to 100 bp. The probes and primers selected (see below) were synthesized by Synthesgen (Houston, TX, USA). Probes were double purified by HPLC to  
20 remove uncoupled dye and evaluated by mass spectroscopy to verify coupling of reporter and quencher dyes to the 5' and 3' ends of the probe, respectively. Their final concentrations were: forward and reverse primers, 900 nM each, and probe, 200nM.

PCR conditions: Normalized RNA from each tissue and each cell line was spotted in each well of a 96 well PCR plate (Perkin Elmer Biosystems). PCR cocktails including  
25 two probes (a probe specific for the target clone and another gene-specific probe multiplexed with the target probe) were set up using 1X TaqMan™ PCR Master Mix for the PE Biosystems 7700, with 5 mM MgCl<sub>2</sub>, dNTPs (dA, G, C, U at 1:1:1:2 ratios), 0.25 U/ml AmpliTaq Gold™ (PE Biosystems), and 0.4 U/ $\mu$ l RNase inhibitor, and 0.25 U/ $\mu$ l reverse transcriptase. Reverse transcription was performed at 48° C for 30 minutes  
30 followed by amplification/PCR cycles as follows: 95° C 10 min, then 40 cycles of 95° C for 15 seconds, 60° C for 1 minute. Results were recorded as CT values (cycle at which a given sample crosses a threshold level of fluorescence) using a log scale, with the difference in RNA concentration between a given sample and the sample with the lowest

CT value being represented as 2 to the power of delta CT. The percent relative expression is then obtained by taking the reciprocal of this RNA difference and multiplying by 100.

In the results for Panel 1, the following abbreviations are used:

- ca. = carcinoma,
- 5 \* = established from metastasis,
- met = metastasis,
- s cell var = small cell variant,
- non-s = non-sm = non-small,
- squam = squamous,
- 10 pl. eff = pl effusion = pleural effusion,
- glio = glioma,
- astro = astrocytoma, and
- neuro = neuroblastoma.

#### 15 **Panel 1.4**

The plates for panel 1.4 include 2 control wells (genomic DNA control and chemistry control) and 94 wells containing cDNA from various samples. The samples in panel 1.4 are broken into 2 classes; samples derived from cultured cell lines and samples derived from primary normal tissues. The cell lines are derived from cancers of the following types: lung cancer, breast cancer, melanoma, colon cancer, prostate cancer, CNS cancer, squamous cell carcinoma, ovarian cancer, liver cancer, renal cancer, gastric cancer and pancreatic cancer. Cell lines used in panel 1.4 are widely available through the American Type Culture Collection, a repository for cultured cell lines. The normal tissues found on panel 1.4 are comprised of pools of samples from 2 to 5 different adult individuals derived from all major organ systems. These samples are derived from the following organs: adult skeletal muscle, fetal skeletal muscle, adult heart, fetal heart, adult kidney, fetal kidney, adult liver, fetal liver, adult lung, fetal lung, various regions of the brain, the spleen, bone marrow, lymph node, pancreas, salivary gland, pituitary gland, adrenal gland, spinal cord, thymus, stomach, small intestine, colon, bladder, trachea, breast, ovary, uterus, placenta, prostate, testis and adipose.

RNA integrity from all samples is controlled for quality by visual assessment of agarose gel electropherograms using 28S and 18S ribosomal RNA staining intensity ratio as a guide (2:1 to 2.5:1 28s:18s) and the absence of low molecular weight RNAs that would

be indicative of degradation products. Samples are controlled against genomic DNA contamination by RTQ PCR reactions run in the absence of reverse transcriptase using probe and primer sets designed to amplify across the span of a single exon.

## 5 **Panel 2**

The plates for Panel 2 generally include 2 control wells and 94 test samples composed of RNA or cDNA isolated from human tissue procured by surgeons working in close cooperation with the National Cancer Institute's Cooperative Human Tissue Network (CHTN) or the National Disease Research Initiative (NDRI). The tissues are derived from  
10 human malignancies and in cases where indicated many malignant tissues have "matched margins" obtained from noncancerous tissue just adjacent to the tumor. These are termed normal adjacent tissues and are denoted "NAT" in the results below. The tumor tissue and the "matched margins" are evaluated by two independent pathologists (the surgical pathologists and again by a pathologists at NDRI or CHTN). This analysis provides a gross  
15 histopathological assessment of tumor differentiation grade. Moreover, most samples include the original surgical pathology report that provides information regarding the clinical stage of the patient. These matched margins are taken from the tissue surrounding (i.e. immediately proximal) to the zone of surgery (designated "NAT", for normal adjacent tissue, in Table RR). In addition, RNA and cDNA samples were obtained from various  
20 human tissues derived from autopsies performed on elderly people or sudden death victims (accidents, etc.). These tissue were ascertained to be free of disease and were purchased from various commercial sources such as Clontech (Palo Alto, CA), Research Genetics, and Invitrogen.

RNA integrity from all samples is controlled for quality by visual assessment of  
25 agarose gel electropherograms using 28S and 18S ribosomal RNA staining intensity ratio as a guide (2:1 to 2.5:1 28s:18s) and the absence of low molecular weight RNAs that would be indicative of degradation products. Samples are controlled against genomic DNA contamination by RTQ PCR reactions run in the absence of reverse transcriptase using probe and primer sets designed to amplify across the span of a single exon.

30

## **Panel 3D**

The plates of Panel 3D are comprised of 94 cDNA samples and two control samples. Specifically, 92 of these samples are derived from cultured human cancer

cell lines, 2 samples of human primary cerebellar tissue and 2 controls. The human cell lines are generally obtained from ATCC (American Type Culture Collection), NCI or the German tumor cell bank and fall into the following tissue groups: Squamous cell carcinoma of the tongue, breast cancer, prostate cancer, melanoma, epidermoid carcinoma, sarcomas, bladder carcinomas, pancreatic cancers, kidney cancers, leukemias/lymphomas, ovarian/uterine/cervical, gastric, colon, lung and CNS cancer cell lines. In addition, there are two independent samples of cerebellum. These cells are all cultured under standard recommended conditions and RNA extracted using the standard procedures. The cell lines in panel 3D and 1.3D are of the most common cell lines used in the scientific literature.

RNA integrity from all samples is controlled for quality by visual assessment of agarose gel electropherograms using 28S and 18S ribosomal RNA staining intensity ratio as a guide (2:1 to 2.5:1 28s:18s) and the absence of low molecular weight RNAs that would be indicative of degradation products. Samples are controlled against genomic DNA contamination by RTQ PCR reactions run in the absence of reverse transcriptase using probe and primer sets designed to amplify across the span of a single exon.

#### **Panel 4**

Panel 4 includes samples on a 96 well plate (2 control wells, 94 test samples) composed of RNA (Panel 4r) or cDNA (Panel 4d) isolated from various human cell lines or tissues related to inflammatory conditions. Total RNA from control normal tissues such as colon and lung (Stratagene, La Jolla, CA) and thymus and kidney (Clontech) were employed. Total RNA from liver tissue from cirrhosis patients and kidney from lupus patients was obtained from BioChain (Biochain Institute, Inc., Hayward, CA). Intestinal tissue for RNA preparation from patients diagnosed as having Crohn's disease and ulcerative colitis was obtained from the National Disease Research Interchange (NDRI) (Philadelphia, PA).

Astrocytes, lung fibroblasts, dermal fibroblasts, coronary artery smooth muscle cells, small airway epithelium, bronchial epithelium, microvascular dermal endothelial cells, microvascular lung endothelial cells, human pulmonary aortic endothelial cells, human umbilical vein endothelial cells were all purchased from Clonetics (Walkersville, MD) and grown in the media supplied for these cell types by Clonetics. These primary cell types were activated with various cytokines or combinations of cytokines for 6 and/or 12-

14 hours, as indicated. The following cytokines were used; IL-1 beta at approximately 1-5 ng/ml, TNF alpha at approximately 5-10 ng/ml, IFN gamma at approximately 20-50 ng/ml, IL-4 at approximately 5-10 ng/ml, IL-9 at approximately 5-10 ng/ml, IL-13 at approximately 5-10 ng/ml. Endothelial cells were sometimes starved for various times by culture in the basal media from Clonetics with 0.1% serum.

Mononuclear cells were prepared from blood of employees at CuraGen Corporation, using Ficoll. LAK cells were prepared from these cells by culture in DMEM 5% FCS (Hyclone), 100  $\mu$ M non essential amino acids (Gibco/Life Technologies, Rockville, MD), 1 mM sodium pyruvate (Gibco), mercaptoethanol  $5.5 \times 10^{-5}$  M (Gibco), and 10 mM Hepes (Gibco) and Interleukin 2 for 4-6 days. Cells were then either activated with 10-20 ng/ml PMA and 1-2  $\mu$ g/ml ionomycin, IL-12 at 5-10 ng/ml, IFN gamma at 20-50 ng/ml and IL-18 at 5-10 ng/ml for 6 hours. In some cases, mononuclear cells were cultured for 4-5 days in DMEM 5% FCS (Hyclone), 100  $\mu$ M non essential amino acids (Gibco), 1 mM sodium pyruvate (Gibco), mercaptoethanol  $5.5 \times 10^{-5}$  M (Gibco), and 10 mM Hepes (Gibco) with PHA (phytohemagglutinin) or PWM (pokeweed mitogen) at approximately 5  $\mu$ g/ml. Samples were taken at 24, 48 and 72 hours for RNA preparation. MLR (mixed lymphocyte reaction) samples were obtained by taking blood from two donors, isolating the mononuclear cells using Ficoll and mixing the isolated mononuclear cells 1:1 at a final concentration of approximately  $2 \times 10^6$  cells/ml in DMEM 5% FCS (Hyclone), 100  $\mu$ M non essential amino acids (Gibco), 1 mM sodium pyruvate (Gibco), mercaptoethanol ( $5.5 \times 10^{-5}$  M) (Gibco), and 10 mM Hepes (Gibco). The MLR was cultured and samples taken at various time points ranging from 1- 7 days for RNA preparation.

Monocytes were isolated from mononuclear cells using CD14 Miltenyi Beads, +ve VS selection columns and a Vario Magnet according to the manufacturer's instructions. Monocytes were differentiated into dendritic cells by culture in DMEM 5% fetal calf serum (FCS) (Hyclone, Logan, UT), 100  $\mu$ M non essential amino acids (Gibco), 1 mM sodium pyruvate (Gibco), mercaptoethanol  $5.5 \times 10^{-5}$  M (Gibco), and 10 mM Hepes (Gibco), 50 ng/ml GM-CSF and 5 ng/ml IL-4 for 5-7 days. Macrophages were prepared by culture of monocytes for 5-7 days in DMEM 5% FCS (Hyclone), 100  $\mu$ M non essential amino acids (Gibco), 1 mM sodium pyruvate (Gibco), mercaptoethanol  $5.5 \times 10^{-5}$  M (Gibco), 10 mM Hepes (Gibco) and 10% AB Human Serum or MCSF at approximately 50 ng/ml. Monocytes, macrophages and dendritic cells were stimulated for 6 and 12-14 hours with

lipopolysaccharide (LPS) at 100 ng/ml. Dendritic cells were also stimulated with anti-CD40 monoclonal antibody (Pharmingen) at 10 µg/ml for 6 and 12-14 hours.

CD4 lymphocytes, CD8 lymphocytes and NK cells were also isolated from mononuclear cells using CD4, CD8 and CD56 Miltenyi beads, positive VS selection columns and a Vario Magnet according to the manufacturer's instructions. CD45RA and CD45RO CD4 lymphocytes were isolated by depleting mononuclear cells of CD8, CD56, CD14 and CD19 cells using CD8, CD56, CD14 and CD19 Miltenyi beads and +ve selection. Then CD45RO beads were used to isolate the CD45RO CD4 lymphocytes with the remaining cells being CD45RA CD4 lymphocytes. CD45RA CD4, CD45RO CD4 and CD8 lymphocytes were placed in DMEM 5% FCS (Hyclone), 100 µM non essential amino acids (Gibco), 1 mM sodium pyruvate (Gibco), mercaptoethanol  $5.5 \times 10^{-5}$  M (Gibco), and 10 mM Hepes (Gibco) and plated at  $10^6$  cells/ml onto Falcon 6 well tissue culture plates that had been coated overnight with 0.5 µg/ml anti-CD28 (Pharmingen) and 3 µg/ml anti-CD3 (OKT3, ATCC) in PBS. After 6 and 24 hours, the cells were harvested for RNA preparation. To prepare chronically activated CD8 lymphocytes, we activated the isolated CD8 lymphocytes for 4 days on anti-CD28 and anti-CD3 coated plates and then harvested the cells and expanded them in DMEM 5% FCS (Hyclone), 100 µM non essential amino acids (Gibco), 1 mM sodium pyruvate (Gibco), mercaptoethanol  $5.5 \times 10^{-5}$  M (Gibco), and 10 mM Hepes (Gibco) and IL-2. The expanded CD8 cells were then activated again with plate bound anti-CD3 and anti-CD28 for 4 days and expanded as before. RNA was isolated 6 and 24 hours after the second activation and after 4 days of the second expansion culture. The isolated NK cells were cultured in DMEM 5% FCS (Hyclone), 100 µM non essential amino acids (Gibco), 1 mM sodium pyruvate (Gibco), mercaptoethanol  $5.5 \times 10^{-5}$  M (Gibco), and 10 mM Hepes (Gibco) and IL-2 for 4-6 days before RNA was prepared.

To obtain B cells, tonsils were procured from NDRI. The tonsil was cut up with sterile dissecting scissors and then passed through a sieve. Tonsil cells were then spun down and resuspended at  $10^6$  cells/ml in DMEM 5% FCS (Hyclone), 100 µM non essential amino acids (Gibco), 1 mM sodium pyruvate (Gibco), mercaptoethanol  $5.5 \times 10^{-5}$  M (Gibco), and 10 mM Hepes (Gibco). To activate the cells, we used PWM at 5 µg/ml or anti-CD40 (Pharmingen) at approximately 10 µg/ml and IL-4 at 5-10 ng/ml. Cells were harvested for RNA preparation at 24, 48 and 72 hours.

To prepare the primary and secondary Th1/Th2 and Tr1 cells, six-well Falcon plates were coated overnight with 10 µg/ml anti-CD28 (Pharmingen) and 2 µg/ml OKT3

(ATCC), and then washed twice with PBS. Umbilical cord blood CD4 lymphocytes (Poietic Systems, German Town, MD) were cultured at  $10^5$  -  $10^6$  cells/ml in DMEM 5% FCS (Hyclone), 100  $\mu$ M non essential amino acids (Gibco), 1 mM sodium pyruvate (Gibco), mercaptoethanol  $5.5 \times 10^{-5}$  M (Gibco), 10 mM Hepes (Gibco) and IL-2 (4 ng/ml). IL-12 (5 ng/ml) and anti-IL4 (1  $\mu$ g/ml) were used to direct to Th1, while IL-4 (5 ng/ml) and anti-IFN gamma (1  $\mu$ g/ml) were used to direct to Th2 and IL-10 at 5 ng/ml was used to direct to Tr1. After 4-5 days, the activated Th1, Th2 and Tr1 lymphocytes were washed once in DMEM and expanded for 4-7 days in DMEM 5% FCS (Hyclone), 100  $\mu$ M non essential amino acids (Gibco), 1 mM sodium pyruvate (Gibco), mercaptoethanol  $5.5 \times 10^{-5}$  M (Gibco), 10 mM Hepes (Gibco) and IL-2 (1 ng/ml). Following this, the activated Th1, Th2 and Tr1 lymphocytes were re-stimulated for 5 days with anti-CD28/OKT3 and cytokines as described above, but with the addition of anti-CD95L (1  $\mu$ g/ml) to prevent apoptosis. After 4-5 days, the Th1, Th2 and Tr1 lymphocytes were washed and then expanded again with IL-2 for 4-7 days. Activated Th1 and Th2 lymphocytes were maintained in this way for a maximum of three cycles. RNA was prepared from primary and secondary Th1, Th2 and Tr1 after 6 and 24 hours following the second and third activations with plate bound anti-CD3 and anti-CD28 mAbs and 4 days into the second and third expansion cultures in Interleukin 2.

The following leukocyte cells lines were obtained from the ATCC: Ramos, EOL-1, KU-812. EOL cells were further differentiated by culture in 0.1 mM dbcAMP at  $5 \times 10^5$  cells/ml for 8 days, changing the media every 3 days and adjusting the cell concentration to  $5 \times 10^5$  cells/ml. For the culture of these cells, we used DMEM or RPMI (as recommended by the ATCC), with the addition of 5% FCS (Hyclone), 100  $\mu$ M non essential amino acids (Gibco), 1 mM sodium pyruvate (Gibco), mercaptoethanol  $5.5 \times 10^{-5}$  M (Gibco), 10 mM Hepes (Gibco). RNA was either prepared from resting cells or cells activated with PMA at 10 ng/ml and ionomycin at 1  $\mu$ g/ml for 6 and 14 hours. Keratinocyte line CCD106 and an airway epithelial tumor line NCI-H292 were also obtained from the ATCC. Both were cultured in DMEM 5% FCS (Hyclone), 100  $\mu$ M non essential amino acids (Gibco), 1 mM sodium pyruvate (Gibco), mercaptoethanol  $5.5 \times 10^{-5}$  M (Gibco), and 10 mM Hepes (Gibco). CCD1106 cells were activated for 6 and 14 hours with approximately 5 ng/ml TNF alpha and 1 ng/ml IL-1 beta, while NCI-H292 cells were activated for 6 and 14 hours with the following cytokines: 5 ng/ml IL-4, 5 ng/ml IL-9, 5 ng/ml IL-13 and 25 ng/ml IFN gamma.

For these cell lines and blood cells, RNA was prepared by lysing approximately  $10^7$  cells/ml using Trizol (Gibco BRL). Briefly, 1/10 volume of bromochloropropane (Molecular Research Corporation) was added to the RNA sample, vortexed and after 10 minutes at room temperature, the tubes were spun at 14,000 rpm in a Sorvall SS34 rotor. The aqueous phase was removed and placed in a 15 ml Falcon Tube. An equal volume of isopropanol was added and left at  $-20$  degrees C overnight. The precipitated RNA was spun down at 9,000 rpm for 15 min in a Sorvall SS34 rotor and washed in 70% ethanol. The pellet was redissolved in 300  $\mu$ l of RNase-free water and 35  $\mu$ l buffer (Promega) 5  $\mu$ l DTT, 7  $\mu$ l RNasin and 8  $\mu$ l DNase were added. The tube was incubated at 37 degrees C for 30 minutes to remove contaminating genomic DNA, extracted once with phenol chloroform and re-precipitated with 1/10 volume of 3 M sodium acetate and 2 volumes of 100% ethanol. The RNA was spun down and placed in RNase free water. RNA was stored at  $-80$  degrees C.

15

#### **Panel 5D and 5I**

The plates for Panel 5D and 5I include two control wells and a variety of cDNAs isolated from human tissues and cell lines with an emphasis on metabolic diseases. Metabolic tissues were obtained from patients enrolled in the Gestational Diabetes study. Cells were obtained during different stages in the differentiation of adipocytes from human mesenchymal stem cells. Human pancreatic islets were also obtained.

In the Gestational Diabetes study subjects are young (18 - 40 years), otherwise healthy women with and without gestational diabetes undergoing routine (elective) Caesarean section. After delivery of the infant, when the surgical incisions were being repaired/closed, the obstetrician removed a small sample ( $<1$  cc) of the exposed metabolic tissues during the closure of each surgical level. The biopsy material was rinsed in sterile saline, blotted and fast frozen within 5 minutes from the time of removal. The tissue was then flash frozen in liquid nitrogen and stored, individually, in sterile screw-top tubes and kept on dry ice for shipment to or to be picked up by CuraGen. The metabolic tissues of interest include uterine wall (smooth muscle), visceral adipose, skeletal muscle (rectus) and subcutaneous adipose. Patient descriptions are as follows:

Patient 2      Diabetic Hispanic, overweight, not on insulin

Patient 7-9    Nondiabetic Caucasian and obese (BMI>30)  
 Patient 10    Diabetic Hispanic, overweight, on insulin  
 Patient 11    Nondiabetic African American and overweight  
 Patient 12    Diabetic Hispanic on insulin

5

Adiocyte differentiation was induced in donor progenitor cells obtained from Osirus (a division of Clonetics/BioWhittaker) in triplicate except for Donor 3U which had only two replicates. Scientists at Clonetics isolated, grew and differentiated human mesenchymal stem cells (HuMSCs) for CuraGen based on the published protocol found in Mark F. Pittenger, et al., Multilineage Potential of Adult Human Mesenchymal Stem Cells *Science* Apr 2 1999: 143-147. Clonetics provided Trizol lysates or frozen pellets suitable for mRNA isolation and ds cDNA production. A general description of each donor is as follows:

15.	Donor 2 and 3: U	Mesenchymal Stem Cells	Undifferentiated
	Donor 2 and 3: AM	Adipose	Adipose Midway Differentiated
	Donor 2 and 3: AD	Adipose	Adipose Differentiated

Human cell lines were generally obtained from ATCC (American Type Culture Collection), NCI or the German tumor cell bank and fall into the following tissue groups: kidney proximal convoluted tubule, uterine smooth muscle cells, small intestine, liver HepG2 cancer cells, heart primary stromal cells, and adrenal cortical adenoma cells. These cells are all cultured under standard recommended conditions and RNA extracted using the standard procedures.

25 All samples were processed at CuraGen to produce single stranded cDNA. RNA integrity from all samples is controlled for quality by visual assessment of agarose gel electropherograms using 28S and 18S ribosomal RNA staining intensity ratio as a guide (2:1 to 2.5:1. 28s:18s) and the absence of low molecular weight RNAs that would be indicative of degradation products. Samples are controlled against genomic DNA contamination by RTQ PCR reactions run in the absence of reverse transcriptase using probe and primer sets designed to amplify across the span of a single exon.

Panel 5I contains all samples previously described with the addition of pancreatic islets from a 58 year old female patient obtained from the Diabetes Research Institute at the

University of Miami School of Medicine. Islet tissue was processed to total RNA at an outside source and delivered to CuraGen for addition to panel 5I.

In the labels employed to identify tissues in the 5D and 5I panels, the following abbreviations are used:

5

GO Adipose = Greater Omentum Adipose

SK = Skeletal Muscle

UT = Uterus

PL = Placenta

10

AD = Adipose Differentiated

AM = Adipose Midway Differentiated

U = Undifferentiated Stem Cells

#### 15 **Panel CNSD.01: Central Nervous System (CNS) Panel**

The plates for Panel CNSD.01 include two control wells and 94 test samples comprised of cDNA isolated from postmortem human brain tissue obtained from the Harvard Brain Tissue Resource Center. Brains are removed from calvaria of donors between 4 and 24 hours after death, sectioned by neuroanatomists, and frozen at -80°C in liquid nitrogen vapor. All brains are sectioned and examined by neuropathologists to confirm diagnoses with clear associated neuropathology.

Disease diagnoses are taken from patient records. The panel contains two brains from each of the following diagnoses: Alzheimer's disease, Parkinson's disease, Huntington's disease, Progressive Supranuclear Palsy, Depression, and "Normal controls". Within each of these brains, the following regions are represented: cingulate gyrus, temporal pole, globus pallidus, substantia nigra, Brodmann Area 4 (primary motor strip), Brodmann Area 7 (parietal cortex), Brodmann Area 9 (prefrontal cortex), and Brodmann area 17 (occipital cortex). Not all brain regions are represented in all cases; e.g., Huntington's disease is characterized in part by neurodegeneration in the globus pallidus, thus this region is impossible to obtain from confirmed Huntington's cases. Likewise Parkinson's disease is characterized by degeneration of the substantia nigra making this region more difficult to obtain. Normal control brains were examined for neuropathology and found to be free of any pathology consistent with neurodegeneration.

RNA integrity from all samples is controlled for quality by visual assessment of agarose gel electropherograms using 28S and 18S ribosomal RNA staining intensity ratio as a guide (2:1 to 2.5:1 28s:18s) and the absence of low molecular weight RNAs that would be indicative of degradation products. Samples are controlled against genomic DNA contamination by RTQ PCR reactions run in the absence of reverse transcriptase using probe and primer sets designed to amplify across the span of a single exon.

In the labels employed to identify tissues in the CNS panel the following abbreviations are used:

10	PSP:	Progressive supranuclear palsy
	Sub Nigra:	Substantia nigra
	Glob Palladus:	Globus pallidus
	Temp Pole:	Temporal pole
	Cing Gyr:	Cingulate gyrus
15	BA:	Brodmann Area

#### **Method of Identifying the Differentially Expressed Gene and Gene Product:**

20 The GeneCalling™ method makes a comparison between experimental samples in the amount of each cDNA fragment generated by digestion with a unique pair of restriction endonucleases, after linker-adaptor ligation, PCR amplification and chromatographic separation. Computer analysis is employed to assign potential identity to the gene fragment. Three methods are routinely used in the identification of a gene fragment found to have altered expression in models of or patients with obesity and/or diabetes.

25 Direct Sequencing: The differentially expressed gene fragment is isolated, cloned into a plasmid and sequenced. Afterwards the sequence information is used to design an oligonucleotide corresponding to either or both termini of the gene fragment. This oligonucleotide, when used in a competitive PCR reaction, will ablate the chromatographic band from which the sequence is derived.

30 Competitive PCR: In competitive PCR, the chromatographic peaks corresponding to the gene fragment of the gene of interest are ablated when a gene-specific primer

(designed from the sequenced band or available databases) competes with primers in the linker-adaptors during the PCR amplification.

5 PCR with Perfect or Mismatched 3' Nucleotides (Trapping): This method utilizes a competitive PCR approach using a degenerate set of primers that extend one or two nucleotides into the gene-specific region of the fragment beyond the flanking restriction sites. As in the competitive PCR approach, primers that lead to the ablation of the chromatographic band add additional sequence information. In conjunction with the size of the gene fragment and the 12 nucleotides of sequence derived from the restriction sites, this additional sequence data can uniquely define the gene after database analysis.

10

#### **Antibodies**

The invention further encompasses antibodies and antibody fragments, such as Fab, (Fab)<sub>2</sub> or single chain FV constructs, that bind immunospecifically to any of the proteins of the invention. Also encompassed within the invention are peptides and polypeptides comprising sequences having high binding affinity for any of the proteins of the invention, including such peptides and polypeptides that are fused to any carrier particle (or biologically expressed on the surface of a carrier) such as a bacteriophage particle.

#### **Methods of Use of the Compositions of the Invention**

20 The protein similarity information, expression pattern, cellular localization, and map location for the protein and nucleic acid disclosed herein suggest that this protein may have important structural and/or physiological functions characteristic of the Ornithine Decarboxylase 1 family. Therefore, the nucleic acids and proteins of the invention are useful in potential diagnostic and therapeutic applications and as a research tool. These include serving as a specific or selective nucleic acid or protein diagnostic and/or prognostic marker, wherein the presence or amount of the nucleic acid or the protein are to be assessed. These also include potential therapeutic applications such as the following: (i) a protein therapeutic, (ii) a small molecule drug target, (iii) an antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), (iv) a nucleic acid useful in gene therapy (gene delivery/gene ablation), (v) an agent promoting tissue regeneration *in vitro* and *in vivo*, and (vi) a biological defense weapon.

The nucleic acids and proteins of the invention have applications in the diagnosis and/or treatment of various diseases and disorders. For example, the compositions of the

present invention will have efficacy for the treatment of patients suffering from: Obesity and/or Diabetes.

These materials are further useful in the generation of antibodies that bind immunospecifically to the substances of the invention for use in diagnostic and/or  
5 therapeutic methods.

#### A. NOV10a - Human Ornithine Decarboxylase 1 – CG124907-01

##### 10 **Discovery Process**

The following sections describe the study design(s) and the techniques used to identify the ornithine decarboxylase 1-gene, encoded protein and any variants, thereof, as being suitable as diagnostic markers, targets for an antibody therapeutic and targets for a small molecule drugs for Obesity and Diabetes.

15

Studies: MB04. Mouse obesity model (genetic)

##### Study Statements:

A large number of mouse strains have been identified that differ in body mass and  
20 composition. The AKR and NZB strains are obese, the SWR, C57L and C57BL/6 strains are of average weight whereas the SM/J and Cast/Ei strains are lean. Understanding the gene expression differences in the major metabolic tissues from these strains will elucidate the pathophysiologic basis for obesity. These specific strains of rat were chosen for differential gene expression analysis because quantitative trait loci (QTL) for body weight  
25 and related traits had been reported in published genetic studies. Tissues included whole brain, skeletal muscle, visceral adipose, and liver.

##### MB.08. Human Mesenchymal Stem Cell differentiation

Bone marrow-derived human mesenchymal stem cells have the capacity to differentiate  
30 into muscle, adipose, cartilage and bone. Culture conditions have been established that permit the differentiation in vitro along the pathway to adipose, cartilage and bone. Understanding the gene expression changes that accompany these distinct differentiation processes would be of considerable biologic value. Regulation of adipocyte differentiation would have importance in the treatment of obesity, diabetes and hypertension. Human

mesenchymal stem cells from 3 donors were obtained and differentiated in vitro according to published methods. RNA from samples of the undifferentiated, mid-way differentiated and fully differentiated cells was isolated for analysis of differential gene expression.

5 BP24.2. Diet induced obesity

The predominant cause for obesity in clinical populations is excess caloric intake. This so-called diet-induced obesity (DIO) is mimicked in animal models by feeding high fat diets of greater than 40% fat content. The DIO study was established to identify the gene expression changes contributing to the development and progression of diet-induced  
10 obesity. In addition, the study design seeks to identify the factors that lead to the ability of certain individuals to resist the effects of a high fat diet and thereby prevent obesity. The sample groups for the study had body weights +1 S.D., + 4 S.D. and + 7 S.D. of the chow-fed controls (below). In addition, the biochemical profile of the + 7 S.D. mice revealed a further stratification of these animals into mice that retained a normal glycemic profile in  
15 spite of obesity and mice that demonstrated hyperglycemia. Tissues examined included hypothalamus, brainstem, liver, retroperitoneal white adipose tissue (WAT), epididymal WAT, brown adipose tissue (BAT), gastrocnemius muscle (fast twitch skeletal muscle) and soleus muscle (slow twitch skeletal muscle). The differential gene expression profiles for these tissues should reveal genes and pathways that can be used as therapeutic targets for  
20 obesity.

Ornithine Decarboxylase 1:

In multiple genecalling studies the enzyme spermidine/spermine acetyl transferase has been found to be dysregulated in various disease models. This enzyme is one of the  
25 rate-limiting enzymes in the production of polyamines spermidine and spermine. Previously, it was shown that oxidation of polyamines leads to generation of hydrogen peroxide, which has been shown to have antilipolytic effect of adipose and may therefore be involved in the progression of obesity. Ornithine decarboxylase catalyzes the first step in polyamine production, which is the conversion of ornithine to putrescine. The polyamine  
30 pathway can be detrimental for the obesity phenotype, since hydrogen peroxide produced during oxidation of polyamines is known to have anti-lipolytic, insulin-like effect on adipocytes. Therefore, inhibiting the production of polyamines and generation of H<sub>2</sub>O<sub>2</sub> by

inhibiting this first enzyme in the polyamine pathway may be beneficial in the treatment for obesity.

The Ornithine Decarboxylase 1 (ODC) is one of the key enzymes in polyamine biosynthesis. Preventing the accumulation of polyamines and their antilipolytic effects by inhibition of ODC at an earlier stage of obesity may inhibit progression of the obesity.

The following is a summary of the findings from the discovery studies, supplementary investigations and assays that also incorporates knowledge in the scientific literature for use of ornithine decarboxylase 1 as a diagnostic and/or target for small molecule drugs and antibody therapeutics.. Taken in total, the data indicates that an inhibitor/antagonist of the human ornithine decarboxylase 1 would be beneficial in the treatment of obesity and/or diabetes.

SPECIES #1 mouse (NZB vs SM/J):

A gene fragment of the mouse spermine/spermidine N-acetyltransferase was initially found to be upregulated by 1.9 fold in the adipose of NZB mice relative to SM/J mice using CuraGen's GeneCalling™ method of differential gene expression. A differentially expressed mouse gene fragment migrating at approximately 411 nucleotides in length (Figure 1a. - red vertical line) was definitively identified as a component of the mouse spermine/spermidine N-acetyltransferase cDNA in NZB and SM/J mouse strains. The method of competitive PCR was used for conformation of the gene assessment. The chromatographic peaks corresponding to the gene fragment of the mouse spermidine/spermine N-acetyltransferase are ablated when a gene-specific primer (see below) which competes with primers in the linker-adaptors during the PCR amplification. The peaks at 411 nt in length are ablated (green trace) in the sample from both the NZB and the SM/J mice. The altered expression in of these genes in the animal model support the role of Ornithine Decarboxylase 1 in the pathogenesis of obesity and/or diabetes.

SPECIES #1 mouse (C57Bl/6 obese euglycemic sd7 vs obese sd1):

A gene fragment of the mouse spermine/spermidine N-acetyltransferase was initially found to be upregulated by 1.8 fold in the epididymal fat pad of the obese euglycemic sd7 mice relative to the obese sd1 mice using CuraGen's GeneCalling™ method of differential gene expression. A differentially expressed rat gene fragment migrating at approximately 178 nucleotides in length (Figure 1a. - red vertical line) was

definitively identified as a component of the mouse spermine/spermidine N-acetyltransferase cDNA in the Troglitazone treated and the untreated SHR control rats. The method of competitive PCR was used for conformation of the gene assessment. The chromatographic peaks corresponding to the gene fragment of the mouse spermidine/spermine N-acetyltransferase are ablated when a gene-specific primer (see below) which competes with primers in the linker-adaptors during the PCR amplification. The peaks at 178 nt in length are ablated (green trace) in the sample from both the C57Bl/6 obese euglycemic sd7 and obese sd1 mice. The altered expression in of these genes in the animal model support the role of Ornithine Decarboxylase 1 in the pathogenesis of obesity and/or diabetes.

#### SPECIES #2 human (adipocyte mid-way vs undifferentiated):

A gene fragment of the human spermine/spermidine N-acetyltransferase was initially found to be upregulated by 1.6 fold in the mid-way human adipocytes relative to the undifferentiated human adipocytes using CuraGen's GeneCalling™ method of differential gene expression. A differentially expressed human gene fragment migrating at approximately 194 nucleotides in length (Figure 1a. - red vertical line) was definitively identified as a component of the human spermine/spermidine N-acetyltransferase cDNA in human mid-way differentiated and undifferentiated adipocytes. The method of competitive PCR was used for conformation of the gene assessment. The chromatographic peaks corresponding to the gene fragment of the human spermine/spermidine N-acetyltransferase are ablated when a gene-specific primer (see below) which competes with primers in the linker-adaptors during the PCR amplification. The peaks at 194 nt in length are ablated (green trace) in the sample from both the human mid-way differentiated and undifferentiated adipocytes. The altered expression of these genes in the human cellular model support the role of Ornithine Decarboxylase 1 in the pathogenesis of obesity and/or diabetes.

Table 1. Spermidine/spermine N-acetyltransferase Gene Sequence identified in NZB vs SM/J mice (Identified fragment from 206 to 616 in **bold**. band size: 411)

```

1 GCTCCCGGGA AACGAATGAG GAACCACTC CTCCTGCTGT TCAAGTACAG GGGCCTGGTG
61 CGCAAAGGGA AGAAAAGCAA AAGACGAAAA TGGCTAAATT TAAGATCCGT CCAGCCACTG
121 CCTCTGACTG CAGTGACATC CTGCGACTGA TCAAGGAACT GGCTAAATAT GAATACATGG
181 AAGATCAAGT CATTTTAACT GAGAAAGATC TCCAAGAGGA TGGCTTTGGA GAACACCCCT
241 TCTACCACTG CCTGGTTGCA GAAGTGCCTA AAGAGCACTG GACCCCTGAA GGACATAGCA
301 TTGTTGGGTT CGCCATGTAC TATTTTACCT ATGACCCATG GATTGGCAAG TTGCTGTATC
361 TTGAAGACTT CTTCTGTATG AGTGATTACA GAGGCTTTGG TATAGGATCA GAAATTTTGA

```

421 AGAATCTAAG CCAGGTTGCC ATGAAGTGTG GCTGCAGCAG TATGCACTTC TTGGTAGCAG  
 481 AATGGAATGA ACCATCTATC AACTTCTACA AAAGAAGAGG TGCTTCGGAT CTGTCCAGTG  
 541 AAGAGGGATG GAGGCTCTTC AAGATTGACA AAGAGTACTT GCTAAAAATG GCAGCAGAGG  
 601 AGTGAGGCGT GCCGGTGTAG ACAATGACAA CCTCCATGTG GCTTTAGAAT AATTCTCAGC  
 5 661 TTCCCTTGCT TTCTATCTTG TGTGTAGTGA AATAATAGAG CGAGCACCCA TTCCAAAGCT  
 721 TTATTACCAG TGACGTTGTT GCATGTTTGA AATTCGGTCT GTTTAAAGTG GCAGTCATGT  
 781 ATGTGGTTTG GAGGCAGAAT TCITGAACAT CTTTGTATGA AGAACCAAGG GGTATGATCT  
 841 TACTATATAA GAAAAACAAA ACITCATTCT TGTGAGTCAT TTAAATGTGT ACAATGTACA  
 10 901 CACTGGTACT TAGAGTTTCT GTTTTGATTG TTTTITTTTA AATAAATCG CTCTTTGATT  
 961 T

Table 2. Spermidine/spermine N-acetyltransferase Gene Sequence identified in C57Bl/6

15 obese euglycemic sd7 vs obese sd1 (Identified fragment from 716 to 893 in **bold**. band size: 178)

235 ACCCCITCTA CCACTGCCTG GTTGCAGAAG TGCCTAAAGA GCACTGGACC CCTGAAGGAC  
 295 ATAGCATTGT TGGGTTCCGC ATGTACTATT TTACCTATGA CCCATGGATT GGCAAGTTGC  
 20 355 TGTATCTTGA AGACTTCTTC GTGATGAGTG ATTACAGAGG CTTTGGTATA GGATCAGAAA  
 415 TTTTGAAGAA TCTAAGCCAG GTTGCCATGA AGTGTGCTG CAGCAGTATG CACTTCTTGG  
 475 TAGCAGAATG GAATGAACCA TCTATCAACT TCTACAAAAG AAGAGGTGCT TCGGATCTGT  
 535 CCAGTGAAGA GGGATGGAGG CTCTTCAAGA TTGACAAAGA GTACTTGCTA AAAATGGCAG  
 595 CAGAGGAGTG AGGCGTGCCG GTGTAGACAA TGACAACCTC CATTGTGCTT TAGAATAATT  
 25 655 CTCAGCTTCC CTGTCTTCT ATCTTGTGTG TAGTGAATA ATAGAGCGAG CACCCATTCC  
 715 AAAGCTTTAT TACCAGTGAC GTTGTGTCAT GTTTGAAATT CGGTCTGTTT AAAGTGGCAG  
 775 TCATGATGTG GGTTTGGAGG CAGAATCTT GAACATCTTT TGATGAAGAA CAAGGTGGTA  
 835 TGATCTTACT ATATAAGAAA AACAAAACCT CATCTTGTG AGTCATTAA ATGTGTACAA  
 30 895 TGTACACACT GGTACTTAGA GTTCTGTTT TGATTCTTTT TTTTAAATA AACTCGCTCT  
 955 TTGATTT

Table 3. Spermidine/spermine N-acetyltransferase Gene Sequence identified in human

35 adipocyte mid-way versus undifferentiated (Identified fragment from 162 to 355 in **bold**. band size: 149).

1 CTGGTGTTTA TCCGTCACTC GCCGAGGTTT CTTGGGTCAT GGTGCCAGCC TGA CTGAGAA  
 61 GAGGACGCTC CCGGGAGACG AATGAGGAAC CACCTCCTCC TACTGTTCAA GTACAGGGGC  
 40 121 CTGGTCCGCA AAGGGAAGAA AAGCAAAAGA CGAAATGGC TAAATTCGTG ATCCGCCCAG  
 181 CCACTGCCGC GACTGTCAGT GACATACTGC GGCTGATCAA GGAGCTGGCT AAATATGAAT  
 241 ACATGGAAGA ACAAGTAATC TTAAGTAAA AAGATCTGCT AGAAGATGGT TTTGGAGAGC  
 301 ACCCCTTTTA CCACTGCCTG GTTGCAGAAG TGCCGAAAGA GCACTGGACT CCGGAAGGTT  
 361 ACAGTCTCTA GCTTCGCCAT GTACATGGCC CTTCCGTGTA CATGGATGGG CGGGGAGGTA  
 45 421 ACTAAAAGAT CCTTTACACA ATAAAGTAGA TGATCATGAT AAATGAGGAC ACAGCATTGT  
 481 TGGTTTTGCC ATGTACTATT TTACCTATGA CCCGTGGATT GGCAAGTTAT TGTATCTTGA  
 541 GGACTTCTTC GTGATGAGTG ATTATAGAGG CTTTGGCATA GGATCAGAAA TTCTGAAGAA  
 601 TCTAAGCCAG GTTGCAATGA GGTGTGCTG CAGCAGCATG CACTTCTTGG TAGCAGAATG  
 661 GAATGAACCA TCCATCAACT TCTATAAAAG AAGAGGTGCT TCTGATCTGT CCAGTGAAGA  
 50 721 GGGTTTGAGA CTGTTCAAGA TCGACAAGGA GTACTTGCTA AAAATGGCAA CAGAGGAGTG  
 781 AGGAGTGCTG CTGTAGATGA CAACCTCCAT TCTATTTTAG AATAAATTC CAACT

Table 4. Human Ornithine Decarboxylase 1 gene and protein sequence.

55 >CG124907-01 1958 nt  
 GCAGGCCAGCCCATGGGGAAGCGCAGACGCCGNGCCTGGGCGCTCTGAGATTGTCACT  
 GCTGTTCCAAGGGCACACGAGAGGATTTGGAATTCCTGGAGAGTTGCCCTTTGTGAGAA  
 GCTGGAATATTCTTTCAATCCATCTCTAGTTTTCCATAGGAACATCAAGAAATCAT

GAACAACTTTGGTAATGAAGAGTTTGACTGCCACTTCCTCGATGAAGGTTTTACTGCCAA  
 GGACATTCCTGGACCAGAAAATTAATGAAGTTCTTCTCTGATGATAAGGATGCCTTCTA  
 TGTGGCAGACCTGGGAGACATTCTAAAGAAACATCTGAGGTGGTTAAAAGCTCTCCCTCG  
 5 TGTCAACCCCTTTTATGCAGTCAAATGTAATGATAGCAAAGCCATCGTGAAGACCCCTTGC  
 TGCTACCGGGACAGGATTTGACTGTGCTAGCAAGACTGAAATACAGTTGGTGACAGTCT  
 GGGGGTGCCTCCAGAGAGGATTATCTATGCAAACTCTGTAAACAAGTATCTCAAATTAA  
 GTATGCTGCTAATAATGGAGTCCAGATGATGACTTTTGATAGTGAAGTTGAGTTGATGAA  
 AGTTGCCAGAGCACATCCCAAAGCAAAGTTGGTTTTCGGGATTGCCACTGATGATTCCAA  
 10 AGCAGTCTGTCGTCTCAGTGTGAAATTCGGTGCACGCTCAGAACAGCAGGCTCCTTTT  
 GGAACGGGCGAAAGAGCTAAATATCGATGTTGTTGGTGTGAGCTTCCATGTAGGAAGCGG  
 CTGTACCGATCCTGAGACCTTCGTGCAGGCAATCTCTGATGCCCGCTGTGTTTGTACAT  
 GGGGGCTGAGGTTGGTTTCAGCATGTATCTGCTTGATATTGGCGGTGGCTTTCCTGGATC  
 TGAGGATGTGAACTTAAATTTGAAGAGATCACCGGCGTAATCAACCAGCGTTGGACAA  
 15 ATACTTTCGCTCAGACTCTGGAGTGAGAATCATAGCTGAGCCCGGAGATACTATGTTGC  
 ATCAGCTTTCACGCTTGCACTTAATATCATTGCCAAGAAAATTGTATTAAAGGAACAGAC  
 GGGCTCTGATGACGAAGATGAGTCGAGTGAGCAGACCTTTATGTATTATGTGAATGATGG  
 CGTCTATGGATCATTAAATGTCATACTCTATGACCAACGACATGATAAGCCCTTCTGCA  
 AAAGAGACCTAAACCAGATGAGAAGTATTATTCATCCAGCATATGGGGACCAACATGTGA  
 20 TGGCCTCGATCGGATTGTTGAGCGCTGTGACCTGCCTGAAATGCATGTGGGTGATTGGAT  
 GCTCTTTGAAAACATGGGCGCTTACACTGTTGCTGCTGCCTCTACGTTCAATGGCTTCCA  
 GAGGCGGACGATCTACTATGTGATGTGAGGCGCTGCGTGGCAACTCATGCAGCAATCCA  
 GAACCCCGACTTCCACCCGAAAGTAGAGGAACAGGATGCCAGCACCTGCTGTGTCTTG  
 TGCCTGGGAGAGTGGGATGAAACGCCACAGAGCAGCCTGTGCTTCGGCTAGTATTAATGT  
 25 GTAGATAGCACTCTGCTAGCTGTAACTGCAAGTTTAGCTTGAATTAAAGGATTTGGGGG  
 GACCATGTAACTTAAATACTGCTAGTTTGAATGTCTTTGTAAGAGTAGGGTCGCCATG  
 ATGCAGCCATATGGAAGACTAGGATATGGGTCACTTATCTGTGTTCTATGGAACATA  
 TTTGAATATTGTTTTATATGGATTTTATTCACTCTTCAGACACGCTACTCAAGAGTGC  
 30 CCTCAGCTGCTGAAACAAGCATTTGTAGCTGTACAATGGCAGAATGGGCCAAAAGCTTA  
 GTGTTGTGACCTGTTTAAATAAAGTATCTTGAATAAACAACAAAAAAGGGGG  
 CCGCCCTAGGGTTCCCAAGTTTACGTACGCTGCATGG

35

Table 5. Human Ornithine Decarboxylase 1 protein sequence.

**ORF Start: 179 ORF Stop: 1562 Frame: 2**

**Human Ornithine Decarboxylase 1 Protein Sequence:**

```

>CG124907-01-prot 461 aa
MNNFGNEEFDCHEFLDEGFTAKDILDQKINEVSSDDKDAFYVADLGDILKKHLRWLKP
RVPFPYAVKCNDSKAI VKTLAATGTGFDCAKTEIQLVQSLGVPPERIIYANPCKQVSQI
KYAANNNGVQMMTFDSEVELMKVARAHEPKALVLR IATDDSKAVCRLSVKFGATLRTSRL
LERAKELNIDVVGVSFHVSGCTDPETFVQAISDARCVFDMGAEVGFMSYLLDIGGFPF
SEDVKLKFEIITGVINPALDKYFPDSGVR IAEPPGRYYVASAFTLAVNI IAKKIVLKEQ
TGSDDDESSEQTFMYVNDGVYGSFNCILYDHAHVKPLQKRPKPDEKYSSSIWGPTC
DGLDRIVERCDLPEMHVGDWMLFENMGAYTVAASTFNGFORPTIYYVMSPAWQLMQQF
QNPDPFPEVEEQDASTLPVSCAWESGMKRHRAACASASINV
  
```

40 Table 6. Clustal W, Protein Domains, Cellular Location and Locus

The following is an alignment of the protein sequences of the human (CG124907-01), rat and mouse versions of the Ornithine Decarboxylase 1.

```

ODC_mouse  MSSTKDEFDCHILDEGFTAKDILDQKINEVSSDDKDAFYVADLGDILKKHLRWLKALP
ODC_rat     MGSFTKEEFDCHILDEGFTAKDILDQKINEVSSDDKDAFYVADLGDVLKKHLRWLKALP
CG124907-01 MNNFGNEEFDCHFLEDEGFTAKDILDQKINEVSSDDKDAFYVADLGDILKKHLRWLKALP

ODC_mouse  RVTFFYAVKCNDSRAIVSTLAAIGTGFDCAKTEIQLVQGLGVPAERVIYANPCKQVSQL
ODC_rat     RVTFFYAVKCNDSRAIVSTLAAIGTGFDCAKTEIQLVQGLGVPPERIIYANPCKQVSQL
CG124907-01 RVTFFYAVKCNDSRAIVKTLAATGTGFDCAKTEIQLVQSLGVPPERIIYANPCKQVSQL

ODC_mouse  KYAASNGVQMMTFDSEIELMKVARAHPKAKLVLRITDDSKAVCRLSVKFGATLKT SRL
ODC_rat     KYAASNGVQMMTFDSEIELMKVARAHPKAKLVLRITDDSKAVCRLSVKFGATLKT SRL
CG124907-01 KYAANNGVQMMTFDSEVELMKVARAHPKAKLVLRITDDSKAVCRLSVKFGATLKT SRL

ODC_mouse  LERAKELNIDVIGVSFHVSGGCTDPETFVQAVSDARCVFDMATEVGFSMHLLDIGGFFPG
ODC_rat     LERAKELNIDVIGVSFHVSGGCTDPETFVQAVSDARCVFDMGTEVGFSMYLLDIGGFFPG
CG124907-01 LERAKELNIDVIGVSFHVSGGCTDPETFVQAVSDARCVFDMGATEVGFSMYLLDIGGFFPG

ODC_mouse  SEDTKLKFEETSVINPALDKYFSPDSGVRIIAEPGRYVVASAFTLAVNIIAKKTIVWKEQ
ODC_rat     SEDTKLKFEETSVINPALDKYFSPDSGVRIIAEPGRYVVASAFTLAVNIIAKKTIVWKEQ
CG124907-01 SEDTKLKFEETSVINPALDKYFSPDSGVRIIAEPGRYVVASAFTLAVNIIAKKTIVWKEQ

ODC_mouse  PGSDDDESN EQTFMYVYVNDGVVGSFNCILYDHAHVKALLQKRPKPDEKYYS SSIWGPTC
ODC_rat     TGSDDDESN EQTLMYVYVNDGVVGSFNCILYDHAHVKALLQKRPKPDEKYYS SSIWGPTC
CG124907-01 TGSDDDESN EQTFMYVYVNDGVVGSFNCILYDHAHVKALLQKRPKPDEKYYS SSIWGPTC

ODC_mouse  DGLDRIVERCNLPEMHVGDWMLFENMGAYTVAAASTFNGFQRPNIYYVMSEPMWQLMHQI
ODC_rat     DGLDRIVERCSLPEMHVGDWMLFENMGAYTVAAASTFNGFQRPNIYYVMSEPMWQLMHQI
CG124907-01 DGLDRIVERCDLPEMHVGDWMLFENMGAYTVAAASTFNGFQRPNIYYVMSEPMWQLMHQI

ODC_mouse  QSHGFPPFEVEEQDDGTLPMSCAQESGMDRHPAACASARINV
ODC_rat     QSHGFPPFEVEEQDVGTLPMSCAQESGMDRHPAACASASINV
CG124907-01 QNPDPFPEVEEQDASTLPVSCAWESGMDRHPAACASASINV

```

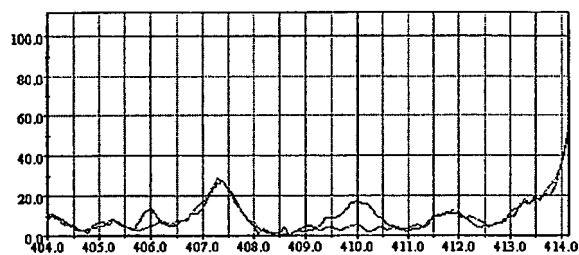
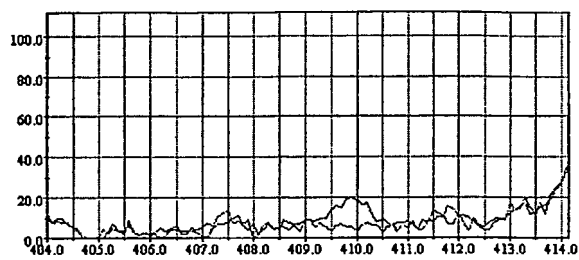
In addition to the human version of the Ornithine Decarboxylase 1 identified as being differentially expressed in the experimental study, other variants have been identified by direct sequencing of cDNAs derived from many different human tissues and from sequences in public databases. No splice-form variants have been identified at CuraGen whereas several amino acid-changing cSNPs were identified. These are found below. The preferred variant of all those identified, to be used for screening purposes, is CG124907-01.

10

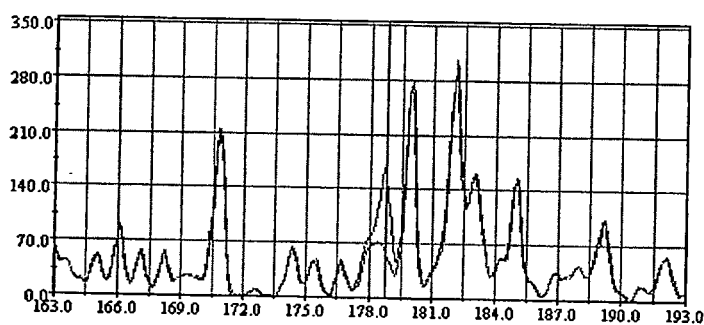
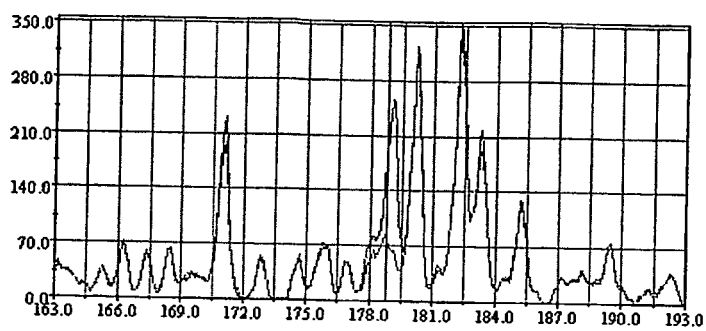
Table 7. Variants of human Ornithine Decarboxylase 1 obtained from direct cloning and/or public databases.

DNA Position	Strand	Alleles	AA Position	AA Change	public SNP #
1447	Minus	C:T	423	Pro => Pro	

15

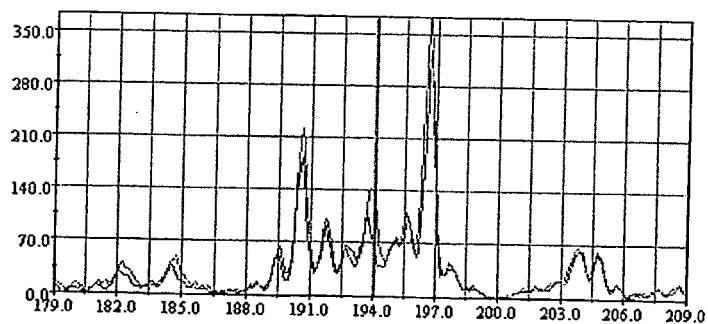


Figures 1A and 1B show differential regulation of spermidine/spermine N-acetyltransferase in the expressed gene fragment in Discovery Study MB.04 of NZB vs SM/J mice. The  
5 abscissa on each graph is measured in length of nucleotides, and the ordinate is measured in signal response.



- Figures 2A and 2B show differential regulation of spermidine/spermine N-acetyltransferase in the expressed gene fragment in Discovery Study MB.04 of NZB vs SM/J mice. The
- 5    abscissa on each graph is measured in length of nucleotides, and the ordinate is measured in signal response.

10



15

Figure 3. Differentially expressed gene fragment in Discovery Study MB.08 identified in human adipocyte mid-way versus undifferentiated, from the human spermidine/spermine N-acetyltransferase.

5

10

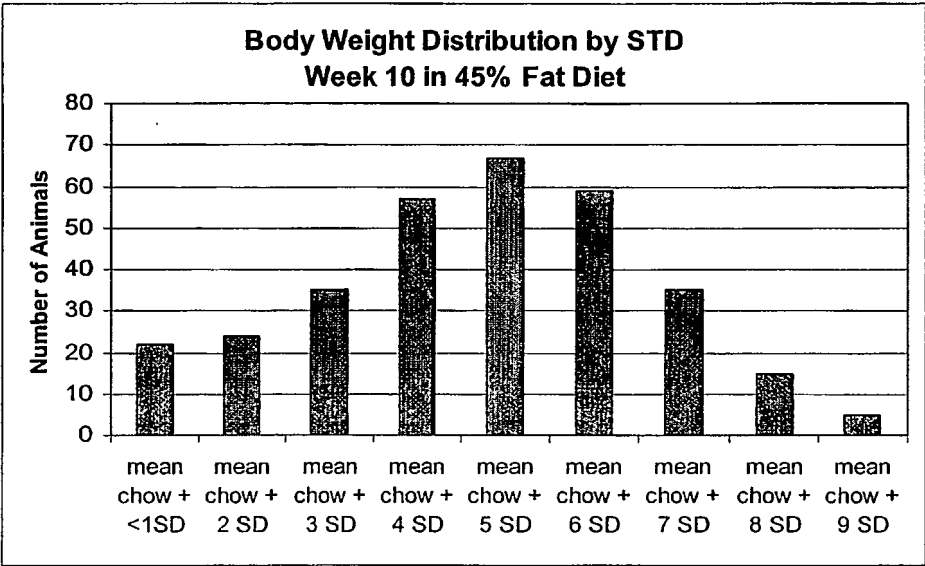


Figure 4. Diet induced obesity Under Discovery Process BP24.2.

15

Species #1 mouse      Strains NZB, SM/J, C56Bl/6  
Species # 2      Human

20

Figure 5 summarize the biochemistry surrounding the human Ornithine Decarboxylase 1 and potential assays that may be used to screen for antibody therapeutics or small molecule drugs to treat obesity and/or diabetes. Cell lines expressing the Ornithine Decarboxylase 1 can be obtained from the RTQ-PCR results shown above.

- 5 These and other Ornithine Decarboxylase 1 expressing cell lines could be used for screening purposes. In the schematic, the biochemistry of "PAO" is that it catalyses oxidation of the secondary amino group of spermine, spermidine and their acetyl derivatives; FAD is the cofactor implicated; and the schematic is shown in monomeric units.

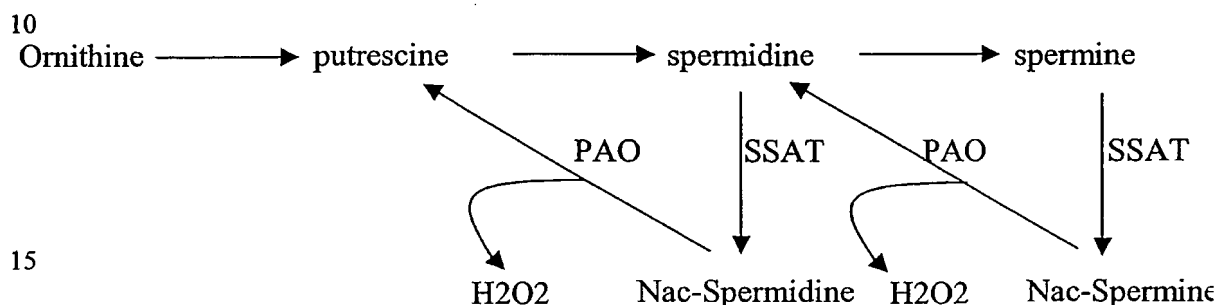
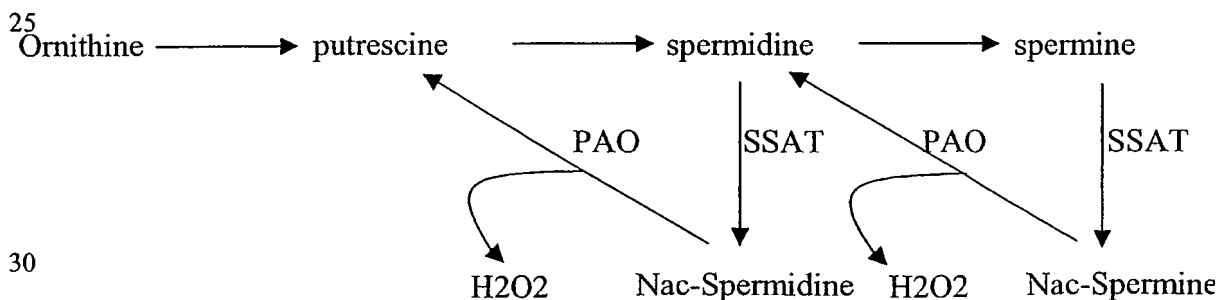
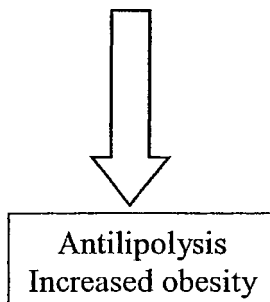


Figure 6 suggests how alterations in expression of the human ornithine decarboxylase 1 and associated gene products function in the etiology and pathogenesis of obesity and/or diabetes. The scheme incorporates the unique findings of these discovery studies in conjunction with what has been reported in the literature. The outcome of inhibiting the action of the human ornithine decarboxylase 1 would be a way to increase lypolysis by inhibiting anti-lypolytic effects of hydrogen peroxide.



5



Ornithine decarboxylase catalyzes the first step in polyamine production, the  
10 conversion of ornithine to putrescine. Inhibiting the production of polyamines and H<sub>2</sub>O<sub>2</sub>  
by inhibiting this first enzyme in the pathway will eliminate the lipolytic effects of H<sub>2</sub>O<sub>2</sub>  
and therefore may be beneficial in the treatment for obesity.

The following is a summary of the findings from the discovery studies,  
supplementary investigations and assays that also incorporates knowledge in the scientific  
15 literature. Taken in total, the data indicates that an inhibitor/antagonist of the human  
Ornithine Decarboxylase 1 would be beneficial in the treatment of obesity and/or diabetes.

In multiple genecalling studies the enzyme spermidine/spermine acetyl transferase  
was found to be dysregulated in various disease models. This enzyme is one of the rate-  
limiting enzymes in the production of polyamines spermidine and spermine. Previously, it  
20 was shown that oxidation of polyamines leads to generation of hydrogen peroxide, which  
has been shown to have antilipolytic effect of adipose and may therefore be involved in the  
progression of obesity. Ornithine decarboxylase catalyzes the first step in polyamine  
production, which is the conversion of ornithine to putrescine. The polyamine pathway can  
be detrimental for the obesity phenotype, since hydrogen peroxide produced during  
25 oxidation of polyamines is known to have anti-lipolytic, insulin-like effect on adipocytes.  
Therefore, inhibiting the production of polyamines and generation of H<sub>2</sub>O<sub>2</sub> by inhibiting  
this first enzyme in the polyamine pathway may be beneficial in the treatment for obesity.

30

**B. NOV12A - Tyrosine aminotransferase – CG135823-01**

The present invention discloses novel associations of proteins and polypeptides and the nucleic acids that encode them with various diseases or pathologies. The proteins and related proteins that are similar to them, are encoded by a cDNA and/or by genomic DNA.

- 5 The proteins, polypeptides and their cognate nucleic acids were identified by CuraGen Corporation in certain cases. The Tyrosine Aminotransferase -encoded protein and any variants, thereof, are suitable as diagnostic markers, targets for an antibody therapeutic and targets for small molecule drugs. As such the current invention embodies the use of recombinantly expressed and/or endogenously expressed protein in various screens to
- 10 identify such therapeutic antibodies and/or therapeutic small molecules.

Table 1. SPECIES #1, Rat Tyrosine Aminotransferase Gene Fragment used for competitive PCR (fragment from 845 to 989 in bold. band size: 145)

15

364 CCTACAGACC CTGAAGTTAC CCAAGCCATG AAAGATGCMC TGGACTCGGG GAAGTACAAT  
 424 GGCTATGCCC CGTCCATCGG CTACCTATCC AGTCGGGAGG AGGTCGCTTC TTACTACCAC  
 484 TGTCAATGAGG CTCCTCTGGA AGCTAAGGAT GTCAATCTGA CAAGCGGCTG CAGTCAGGCC  
 544 ATTGAGCTAT GTCTAGCTGT GTTGGCCAAT CCTGGACAAA ACATCCTCAT TCCAAGGCCC  
 20 604 GGGTTTTCCT TCTATAGGAC TTTGGCTGAG TCTATGGGAA TTGAGGTCAA GCTCTACAAT  
 664 CTCCTGCCCC AGAAGTCTTG GGAATTGAC CTAAACAAC TGAATCTCT GATCGATGAA  
 724 AAAACAGCGT GTCTTGTGT CAACAACCA TCCAATCCCT GTGGCTCCGT GTTCAGTAAG  
 784 CGACACCTTC AGAAGATTTT GGCAGTGGCT GAAAGGCAGT GTGTCCCAT CTTAGCTGAC  
 844 GAGATCTATG GTGACATGGT GTTTTCAGAT TGCAAATACG AACCCTGGC CAACCTCAGC  
 25 904 ACCAATGTTT CCATCCTGTC CTGTGGTGGG CTGGCCAAGC GCTGGCTGGT CCTTGGCTGG

964 AGGTTGGGCT GGATCCTCAT TCATGATCGA AGAGACATTT TTGGCAATGA GATTCGAGAC  
 1024 GGGCTGGTGA AACTGAGTCA GCGGATCCTG GGACCATGCA CCATAGTCCA GGGTGCTCTG  
 1084 AAGAGCATCC TTCAGCGAAC CCCTCAGGAG TTCTATCACG ACACGTTAAG CTTCCCTCAAG  
 30 1144 TCCAATGCGG ACCTCTGCTA TGGGGCACTG GCTGCCATCC CTGGACTCCA GCCGGTCCGC  
 1204 CCTTCTGGAG CCATGTACCT TATGGTGGGA ATTGAGATGG AGCATTTCCT GGAATTCGAG  
 1264 AACGACGTGG AGTTCACAGA GCGGTTGATT GCGGAGCAGG CTGTCCACTG TCTCCAGCA  
 1324 ACGTGCTTCG AGTACCCAAA TTTCTTCCGA GTGGTCATCA CAGTCCCCGA GGTGATGATG  
 1384 CTGGAGGCTT GTAGCCGGAT CCAGGAGTTC TGTGAACAGC ACTACCACTG TGCTGAAGGC  
 35 1444 AGCCAGGAGG AGTGTGACAA ATAAGC

(gene length is 2364, only region from 364 to 1469 shown)

Table 2. SPECIES #2, Rat Tyrosine Aminotransferase Gene Fragment used for competitive PCR (fragment from 1 to 277 in bold. band size: 277).

40

1 TCATGATCGA AGAGACGTTT TTGGCAATGA GATTCGAGAC GGGCTGGTGA AACTGAGTCA  
 61 GCGGATCCTG GGACCATGCA CCATAGTCCA GGGTGCTCTG AAGAGCATCC TTCAGCGAAC

(gene length is 277, only region from 1 to 277 shown)

Table 3. SPECIES #3, Mouse Tyrosine Aminotransferase Gene Fragment used for competitive PCR (fragment from 57 to 275 in **bold**, band size: 220)

1 CCTTCAGAAG ATTTTGGCAG TGGCTGAAAG GCAATGCGTC CCCATCTTAG CCGATGAGAT  
61 CTATGTGTGAC ATGGTGTGTT CAGATTGCAAA ATAGGAACCA CTCGACACCC TCAGACACCA  
121 TGTGCCCATC CTGTCTGTGT TGTGGCTGGC CATCGCTGTG GTTGTCCTCC GCTGGAGGCT  
181 GGGCTGGATC CTTATCCATG ATCGAAGAGA CATTTTTGGC AATGAGATTG GGGACGGGCT  
241 GTTGAAGCTG AGTCACGCGA TCGTGGCCCC GTGCACCATC GTCCAGGTTG CCTCGAAGAG  
301 CATCTTTCAG CGCACCCCTC AGGATTTCTA CACGAGAACT TTAAGCTTCC TTAAGTCCAA  
361 TGGCGACCTC TGCTATGGGG CGTTGTCTGC AATTCTTGGA CTCGACCGCC TCGGCCATC  
421 TGGAGCATG TACCTATTGG TGGGAATTGA GATGGAGCAC TTCCAGAAAT TTGAGAATGA  
481 CGTGGAAATC ACAGAGCGGT TAATTGCGCG AGNNTCTGTC GNACTGCTCC AGCAGTGCTC  
541 TCGAGTACCA ATTTCTTCCG GGTGTCATAC AGTCCCGAG TGTATGATCT G

**Table 4. Human Tyrosine Aminotransferase gene and protein sequence.**

736

TATATTATCTTTTCATACATTTTCTAAGAAACATTATATTGATAAGATCTTTTATTTTG  
CAAGGCATAAATTATTGTTTTTCTTTTTTTTTTTAATAAATTTACCAAGT

5 Table 5. Amino Acid sequence of Human Aminotransferase.

ORF Start: 97 ORF Stop: 1459 Frame: 1

Human Tyrosine Aminotransferase Protein Sequence:

>CG135823-01-prot 454 aa

MDPYMIQMSSKGNLPSILDVHVNVGGRSSVPGMKGRKARWSVRPSDMAKKTFFNPRAIV  
DNMKVKPNPNKTMISLSIGDPTVFGNLPDPEVTQAMKDALDSGKYNGYAPSIGLSSRE  
ETASYHCPEAPLEAKDVILTSGCSQAIDLCLAVLANPGQNILVPRPGFSLYKTLAESMG  
IEVKLYNLLPEKSWEIDLKQLEYLIDEKTACLI VNNPSNPCGSVFSKRHLQKILAVARQ  
CVPILADEIYGDMVFSCKYEPLATLSTDVPILSCGGLAKRWLVPGWRLGWILIHDRDI  
FGNEIRDGLVKLSQRILGPCTIVQGALKSILCRTPGEFYHNTLSFLKSNADLCYGALAAI  
PGLRPVRPSGAMVLMVGIEMEHFPEFENDVEFTERLVAEQSVHCLPATCFEYPNFI R VVI  
TVPEVMMLEACSRIQEFCEQHYHCAEGSQEECDK

10.

Table 6. Clustal W, Protein Domains, Cellular Location and Locus

The following is an alignment of the protein sequences of the human (CG135823-01) and rat versions of the Tyrosine Aminotransferase.

15

TAT_rat	MDSVYIQT	TDVDD S	LSSVLDVHVN	VGGRSSV	PGMKGRKARWD	VRPSDMSN	KTFNPRAIV
CG135823-01	MDPYMIQMSSKGN	LPSILDVHVN	VGGRSSV	PGMKGRKARWS	VRPSDMAK	KTFNPRAIV	
TAT_rat	DNMKVQPNPNKT	ISLSIGDPTV	FGNLPDPEVT	QAMKDALDSGKYNGYAPSIGLSSRE			
CG135823-01	DNMKVKPNPNKTM	ISLSIGDPTVFGNLPDPEVT	QAMKDALDSGKYNGYAPSIGLSSRE				
TAT_rat	EVASVYHC	PEAPLEAKDVILTSGCSQAIDLCLAVLANPGQNILVPRPGFSLYKTLAESMG					
CG135823-01	ETASVYHC	PEAPLEAKDVILTSGCSQAIDLCLAVLANPGQNILVPRPGFSLYKTLAESMG					
TAT_rat	IEVKLYNLLPEKSWEIDLKQLES	LIDEKTACLI	VNNPSNPCGSVFSKRHLQKILAVARQ				
CG135823-01	IEVKLYNLLPEKSWEIDLKQLEYLIDEKTACLI	VNNPSNPCGSVFSKRHLQKILAVARQ					
TAT_rat	CVPILADEIYGDMVFSCKYEPLATLSTDVPILSCGGLAKRWLVPGWRLGWILIHDRDI						
CG135823-01	CVPILADEIYGDMVFSCKYEPLATLSTDVPILSCGGLAKRWLVPGWRLGWILIHDRDI						
TAT_rat	FGNEIRDGLVKLSQRILGPCTIVQGALKSILCRTPGEFYHNTLSFLKSNADLCYGALAAI						
CG135823-01	FGNEIRDGLVKLSQRILGPCTIVQGALKSILCRTPGEFYHNTLSFLKSNADLCYGALAAI						
TAT_rat	PGLRQPVVRPSGAMVLMVGIEMEHFPEFENDVEFTERLVAEQSVHCLPATCFEYPNFI R VVI						
CG135823-01	PGLRQPVVRPSGAMVLMVGIEMEHFPEFENDVEFTERLVAEQSVHCLPATCFEYPNFI R VVI						
TAT_rat	TVPEVMMLEACSRIQEFCEQHYHCAEGSQEECDK						
CG135823-01	TVPEVMMLEACSRIQEFCEQHYHCAEGSQEECDK						

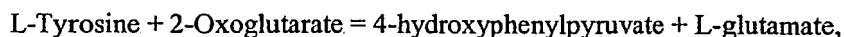
20 Human Tyrosine Aminotransferase:

Locus: 16q22.1 (QTL for intracellular fat on 16q22)

Intracellular

## Biochemistry and Cell Line Expression

Tyrosine Aminotransferase catalyses the following reaction:



5. using pyridoxal 5'-phosphate as a cofactor.

- Tyrosine Aminotransferase activity was measured usually by fix-time assay (measurement of tyrosine absorbance by spectrophotometry). Liver extract, primary hepatocytes and different hepatocyte cell lines were reported to utilize as a source of TAT.
- 10 Cell lines expressing the Tyrosine Aminotransferase can be obtained from the RTQ-PCR results shown above. These and other Tyrosine Aminotransferase expressing cell lines could be used for screening purposes.

- In addition to the human version of the Tyrosine Aminotransferase identified as being differentially expressed in the experimental study, other variants have been identified
- 15 by direct sequencing of cDNAs derived from many different human tissues and from sequences in public databases. No splice-form variants have been identified at CuraGen whereas several amino acid-changing cSNPs were identified in literature. Described below SNPs cause activity deficiency of TAT and were associated with disease called tyrosinemia, type II.

20

Natt E, Kida K, Odievre M, Di Rocco M, Scherer G.

Point mutations in the tyrosine aminotransferase gene in tyrosinemia type II.

Proc. Natl. Acad. Sci. U S A 1992 Oct 1;89(19):9297-301.

PMID: 1357662

25

Table 7. Variants of the human Tyrosine Aminotransferase obtained from direct cloning and/or public databases.

DNA Position	Strand	Alleles	AA Position	AA Change	public SNP #
223		C:G	74	Ser → Stop	
1086		G:T	417	Arg → Stop	
1251		G:T	362	Gly → Val	

30

There are several reasons to use tyrosine aminotransferase as a diagnostic and/or target for small molecule drugs and antibody therapeutics.:

1. Tyrosine Aminotransferase is a rate-limiting enzyme in phenylalanine/tyrosine catabolism, which may contribute to gluconeogenesis and lipid biosynthesis. The level of enzyme is induced by glucocorticoids, and the excess of glucocorticoids frequently results in obesity, insulin resistance and glucose intolerance.
2. Up-regulation of TAT in MB.05 study may contribute to insulin resistance in HTG rats, in MB.01 - to hyperglycemia in SHR rats. Down-regulation of TAT in response to troglitazone treatment in MB.01 study suggests that TAT may be one of downstream targets for this antidiabetic drug.
3. On the other hand, down-regulation of TAT in BP24.02 study may represent the compensatory mechanism to decrease lipid biosynthesis in obese animals.
4. Taken in total, the data indicates that an inhibitor of the human Tyrosine Aminotransferase would be beneficial in the treatment of obesity.

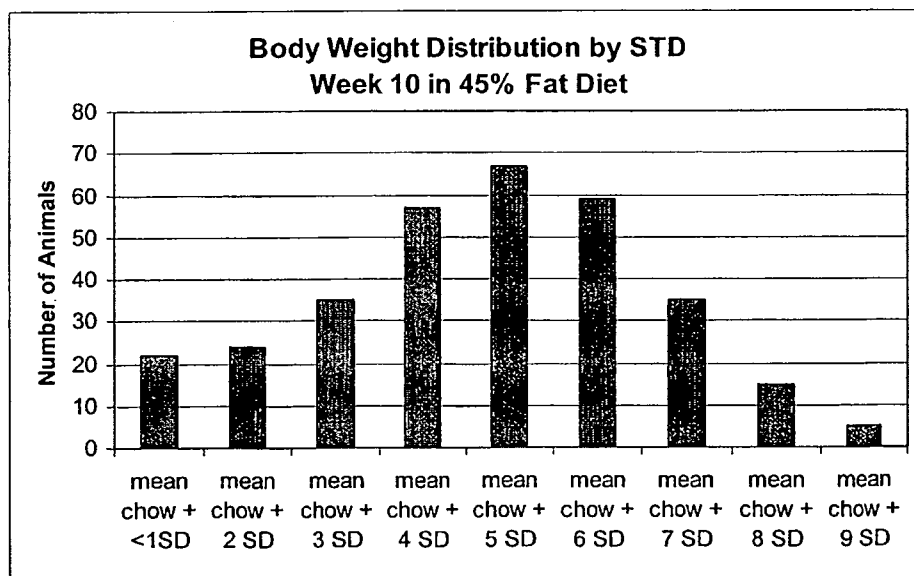


Figure 1. Bar Graph of Diet induced obesity Under Discovery Process BP24.2.

Species #1	Rat	Strains	HTG, Lewis, Wistar
Species #2	Rat	Strains	SHR, SD
Species #3	Mouse	Strains	C57BL/6J

5

Figures 2A, 2B, 2C, 2D, 2E, and 2F. Differentially expressed gene fragments in rat (SPECIES #1); rat (SPECIES #2) and mouse (SPECIES #3) Tyrosine Aminotransferase.

SPECIES #1. Figures 2A and 2B show differentially expressed gene fragments in  
 10 Discovery Study MB.05. from the rat tyrosine aminotransferase (in the graphs, the abscissa is measured in lengths of nucleotides and the ordinate is measured as a signal response). A gene fragment of the rat Tyrosine Aminotransferase was initially found to be up-regulated by 1.7 fold in the muscle and liver tissues of HTG rat relative to normal control rat strain using CuraGen's GeneCalling™ method of differential gene expression. A differentially  
 15 expressed rat gene fragment migrating, at approximately 145 nucleotides in length (Figure 2A - red vertical line) was definitively identified as a component of the rat Tyrosine Aminotransferase cDNA. The method of competitive PCR was used for conformation of the gene assessment. The electropherogramatic peaks corresponding to the gene fragment of the rat Tyrosine Aminotransferase are ablated when a gene-specific primer (see below)  
 20 competes with primers in the linker-adaptors during the PCR amplification. The peaks at 145 nt in length are ablated (green trace) in the sample from both the HTG and control rats.

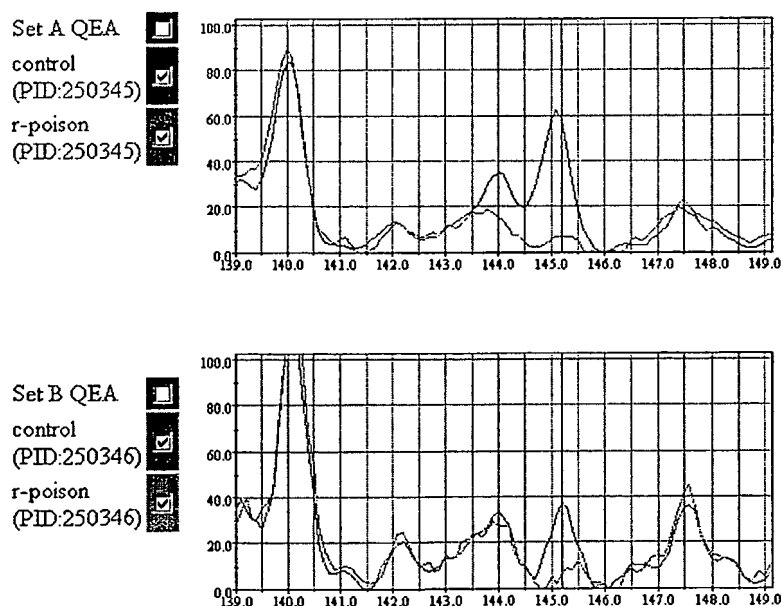
SPECIES #2. Figures 2C and 2D show differentially expressed gene fragments in  
 25 Discovery Study MB.01. from rat tyrosine aminotransferase (in the graphs, the abscissa is measured in lengths of nucleotides and the ordinate is measured as a signal response). The gene fragments corresponding to the rat TAT were found to be up-regulated in liver tissues of SHR rat relative to normal control rat strain, and to be down-regulated in the liver of SHR rat in response to troglitazone treatment. A differentially expressed rat gene fragment migrating, at approximately 277.4 nucleotides in length (Figure 2C - red vertical line) was  
 30 definitively identified as a component of the rat Tyrosine Aminotransferase cDNA by the method of competitive PCR. The electropherogramatic peaks corresponding to the gene fragment of the rat Tyrosine Aminotransferase are ablated when a gene-specific primer (see below) competes with primers in the linker-adaptors during the PCR amplification. The

peaks at 277.4 nt in length are ablated (green trace) in the sample from both the SHR rat liver treated and untreated with troglitazone.

- SPECIES #3 Figures 2E and 2F show differentially expressed gene fragments in
- 5 Discovery Study BP24.02 from mouse tyrosine aminotransferase (in the graphs, the abscissa is measured in lengths of nucleotides and the ordinate is measured as a signal response). Additionally, gene fragments corresponding to the mouse TAT were found to be down-regulated in liver tissues of hyperglycemic fat mouse (hgsd7) relative to normal animal on low fat diet (chow) in a mouse model of dietary-induced obesity. A differentially
- 10 expressed mouse gene fragment migrating, at approximately 220.3 nucleotides in length (Figure 2A - red vertical line) was definitively identified as a component of the mouse Tyrosine Aminotransferase cDNA by the method of competitive PCR. The chromatographic peaks corresponding to the gene fragment of the mouse Tyrosine
- 15 Aminotransferase are ablated when a gene-specific primer (see below) competes with primers in the linker-adaptors during the PCR amplification in the sample from both the hyperglycemic fat mouse relative and normal animals. The altered expression in of these genes in the animal model support the role of the Tyrosine Aminotransferase in the pathogenesis of obesity and/or diabetes.

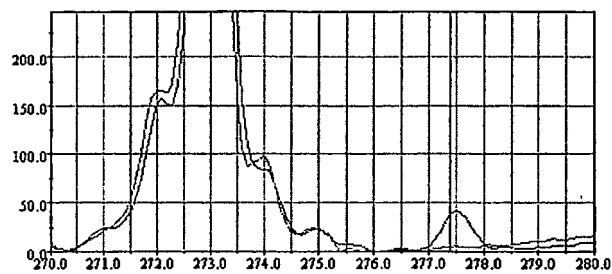
20

## SPECIES #1

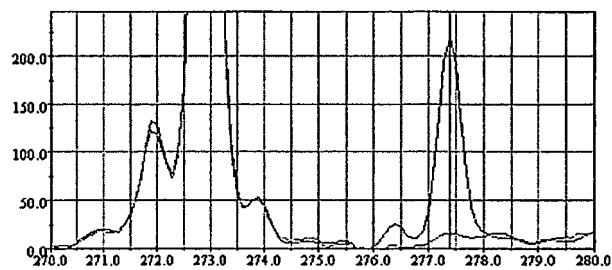


SPECIES #2

Set A QEA  
control  
(PID:99585)  
r-poison  
(PID:99585)



Set B QEA  
control  
(PID:99584)  
r-poison  
(PID:99584)



5

10

SPECIES #3

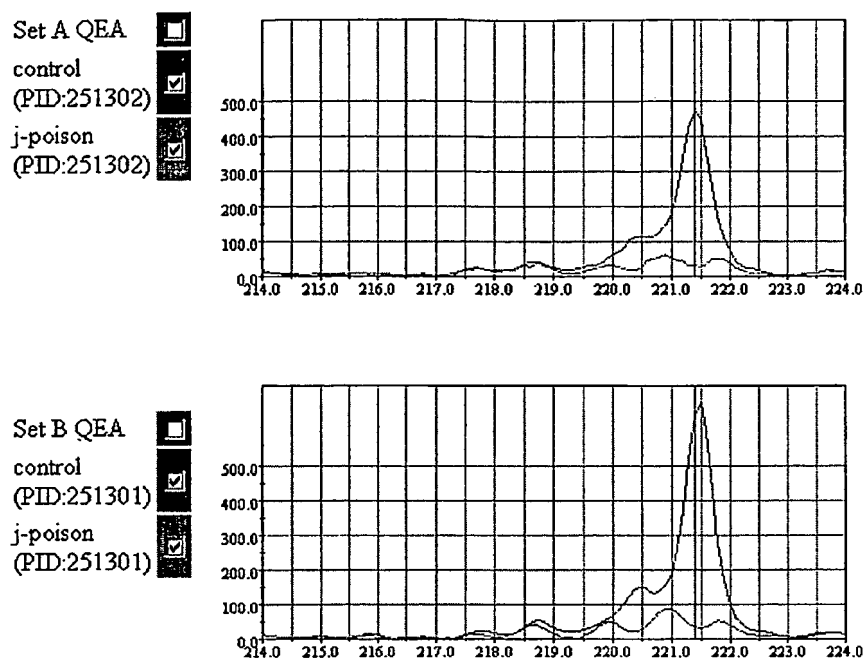


Figure 3. SAGE Data Results

5	<u>SAGE Duke 1273</u>	128		5	38836
	<u>SAGE Duke H1020</u>	76		4	52371
	<u>SAGE HCT116</u>	116		7	60322
	<u>SAGE CAPAN1</u>	158		6	37926
10	<u>SAGE OV1063-3</u>	128		5	38938
	<u>SAGE Tu102</u>	121		7	57636
	<u>SAGE 293-IND</u>	122		3	24481

15

20

25

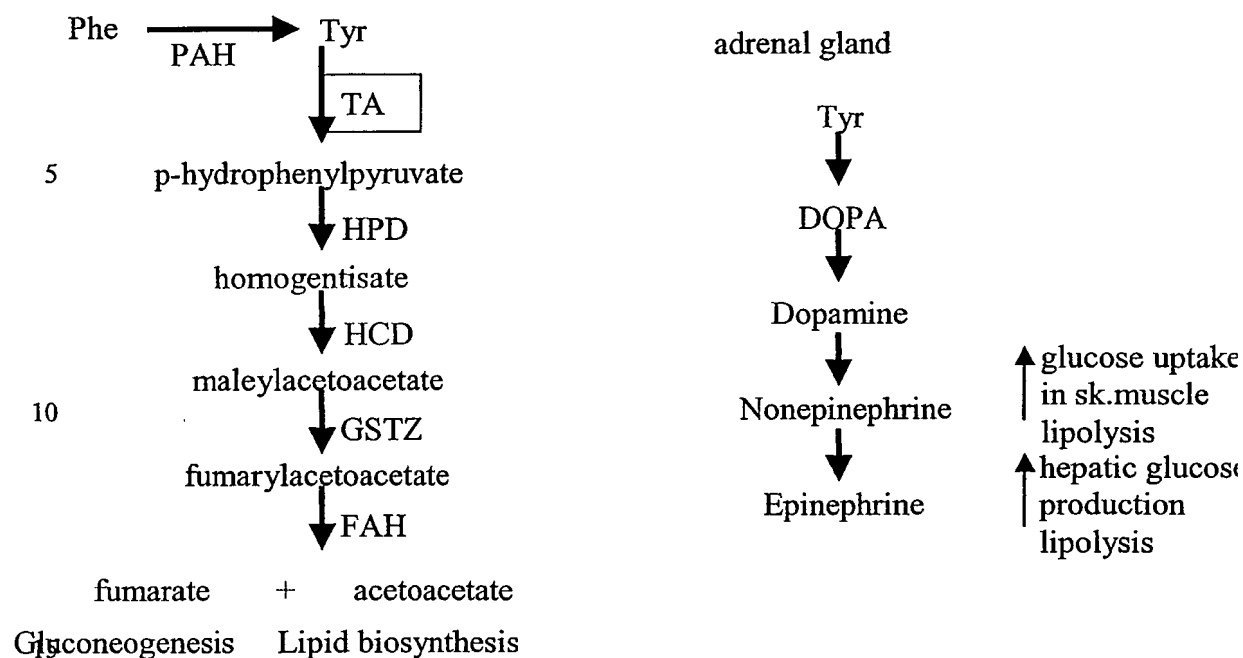


Figure 4 shows pathways that are relevant to the etiology and pathogenesis of obesity and/or diabetes. This figure illustrates the catabolism of tyrosine and phenylalanine and suggests how alterations in expression of the human Tyrosine Aminotransferase and associated gene products function in the etiology and pathogenesis of obesity and/or diabetes. The scheme incorporates the unique findings of these discovery studies in conjunction with what has been reported in the literature. The outcome of inhibiting the action of the human Tyrosine Aminotransferase would inhibit the contribution of these catabolic pathways to gluconeogenesis and lipid biosynthesis and would be beneficial for the treatment of obesity and/or diabetes.

**C. NOV13A - Human Polyamine oxidase – CG140122-01**

The present invention discloses novel associations of proteins and polypeptides and the nucleic acids that encode them with various diseases or pathologies. The proteins and related proteins that are similar to them, are encoded by a cDNA and/or by genomic DNA.

- 5 The proteins, polypeptides and their cognate nucleic acids were identified by CuraGen Corporation in certain cases. The Polyamine Oxidase -encoded protein and any variants, thereof, are suitable as diagnostic markers, targets for an antibody therapeutic and targets for small molecule drugs. As such the current invention embodies the use of recombinantly expressed and/or endogenously expressed protein in various screens to identify such
- 10 therapeutic antibodies and/or therapeutic small molecules.

**Discovery Process**

- The following sections describe the study design(s) and the techniques used to
- 15 identify the Polyamine oxidase-encoded protein and any variants, thereof, as being suitable as diagnostic markers, targets for an antibody therapeutic and targets for a small molecule drugs for Obesity and Diabetes.

Studies: MB04. Mouse obesity model (genetic)

- 20 Study Statements:

- A large number of mouse strains have been identified that differ in body mass and composition. The AKR and NZB strains are obese, the SWR, C57L and C57BL/6 strains are of average weight whereas the SM/J and Cast/Ei strains are lean. Understanding the gene expression differences in the major metabolic tissues from these strains will elucidate
- 25 the pathophysiologic basis for obesity. These specific strains of rat were chosen for differential gene expression analysis because quantitative trait loci (QTL) for body weight and related traits had been reported in published genetic studies. Tissues included whole brain, skeletal muscle, visceral adipose, and liver.

- 30 MB.08. Human Mesenchymal Stem Cell differentiation

Bone marrow-derived human mesenchymal stem cells have the capacity to differentiate into muscle, adipose, cartilage and bone. Culture conditions have been established that permit the differentiation in vitro along the pathway to adipose, cartilage and bone. Understanding the gene expression changes that accompany these distinct differentiation

processes would be of considerable biologic value. Regulation of adipocyte differentiation would have importance in the treatment of obesity, diabetes and hypertension. Human mesenchymal stem cells from 3 donors were obtained and differentiated in vitro according to published methods. RNA from samples of the undifferentiated, mid-way differentiated and fully differentiated cells was isolated for analysis of differential gene expression.

#### BP24.2. Diet induced obesity

The predominant cause for obesity in clinical populations is excess caloric intake. This so-called diet-induced obesity (DIO) is mimicked in animal models by feeding high fat diets of greater than 40% fat content. The DIO study was established to identify the gene expression changes contributing to the development and progression of diet-induced obesity. In addition, the study design seeks to identify the factors that lead to the ability of certain individuals to resist the effects of a high fat diet and thereby prevent obesity. The sample groups for the study had body weights +1 S.D., +4 S.D. and +7 S.D. of the chow-fed controls (below). In addition, the biochemical profile of the +7 S.D. mice revealed a further stratification of these animals into mice that retained a normal glycemic profile in spite of obesity and mice that demonstrated hyperglycemia. Tissues examined included hypothalamus, brainstem, liver, retroperitoneal white adipose tissue (WAT), epididymal WAT, brown adipose tissue (BAT), gastrocnemius muscle (fast twitch skeletal muscle) and soleus muscle (slow twitch skeletal muscle). The differential gene expression profiles for these tissues should reveal genes and pathways that can be used as therapeutic targets for obesity. The bar graph in Figure 1 indicates results.

#### Polyamine oxidase:

In multiple genecalling studies we have found the enzyme spermidine/spermine acetyl transferase to be dysregulated in various disease models (see below). This enzyme is one of the rate-limiting enzymes in the production of polyamines spermidine and spermine (see Figure 6). Figure 6 shows pathways where alterations in expression of the human polyamine oxidase and associated gene products function in the etiology and pathogenesis of obesity and/or diabetes. The scheme incorporates the unique findings of these discovery studies in conjunction with what has been reported in the literature. The outcome of inhibiting the action of the human polyamine oxidase would be a way to increase lipolysis by inhibiting anti-lipolytic effects of hydrogen peroxide..

Previously, it was shown that oxidation of polyamines leads to generation of hydrogen peroxide, which has been shown to have antilipolytic effect of adipose and may therefore be involved in the progression of obesity. The enzyme catalyzing the reaction where hydrogen peroxide is produced, i.e. oxidation of secondary amino group of spermine, spermidine and their acetyl derivatives, is polyamine oxidase. Therefore, we nominate the enzyme polyamine oxidase as a valuable tool to inhibit the polyamine pathway and the production of hydrogen peroxide.

#### 10 **Rationale for use as a diagnostic and/or target for small molecule drugs and antibody therapeutics:**

The following is a summary of the findings from the discovery studies, supplementary investigations and assays that also incorporates knowledge in the scientific literature. Taken in total, the data indicates that an inhibitor/antagonist of the human Polyamine oxidase would be beneficial in the treatment of obesity and/or diabetes (Figure 5 shows biochemistry for human polyamine oxidase and assays that may be used to screen for antibody therapeutics or small molecule drugs to treat obesity and/or diabetes. Cell lines expressing the polyamine oxidase can be obtained from the RTQ-PCR results shown above. These and other polyamine oxidase-expressing cell lines could be used for screening purposes.

Table 1. Spermidine/spermine N-acetyltransferase Gene Sequence identified in NZB vs SM/J mice

(Identified fragment from 206 to 616 in **bold**. band size: 411)

```

25      1 GCTCCCGGGA AACGAATGAG GAACCACTC CTCCTGCTGT TCAAGTACAG GGGCCTGGTG
      61 CGCAAAGGGA AGAAAAGCAA AAGACGAAAA TGGCTAAATT TAAGATCCGT CCAGCCACTG
      121 CCTCTGACTG CAGTGACATC CTGCGACTGA TCAAGGAACT GGCTAAATAT GAATACATGG
      181 AAGATCAAGT CATTITAACT GAGAAAGATC TCCAAGAGGA TGGCTTTGGA GAACACCCCT
      241 TCTACCACTG CCTGGTTGCA GAAGTGCCTA AAGAGCACTG GACCCCTGAA GGACATAGCA
      301 TTGTTGGGTT CGCCATGTAC TATTTTACCT ATGACCCATG GATTGGCAAG TTGCTGTATC
      361 TTGAAGACTT CTTCGTGATG AGTGATTACA GAGGCTTTGG TATAGGATCA GAAATTTTGA
      421 AGAATCTAAG CCAGGTTGCC ATGAAGTGTC GCTGCAGCAG TATGCACTTC TTGGTAGCAG
      481 AATGGAATGA ACCATCTATC AACTTCTACA AAAGAAGAGG TGCTTCGGAT CTGTCCAGTG
      541 AAGAGGGATG GAGGCTCTTC AAGATTGACA AAGAGTACTT GCTAAAAATG GCAGCAGAGG
      601 AGTGAGGCGT GCCGGTGTAG ACAATGACAA CCTCCATTGT GCTTTAGAAAT AATCTCAGC
      661 TTCCCTTGCT TTCTATCTTG TGTGTAGTGA AATAATAGAG CGAGCACCCA TTCCAAAGCT
      721 TTATTACCAG TGACGTTGTT GCATGTTTGA AATTCGGTCT GTTTAAAGTG GCAGTCATGT
      781 ATGTGGTTTG GAGGCAGAAAT TCITGAACAT CTTTGTATGA AGAACAAAGT GGTATGATCT
      841 TACTATATAA GAAAAACAAA ACTTCATTCT TGTGAGTCAT TTAATGTGT ACAATGTACA
      901 CACTGGTACT TAGAGTTTCT GTTTTGATTC TTTTITTTTA AATAAACTCG CTCPTTGATT
      961 T

```

Table 2. Spermidine/spermine N-acetyltransferase Gene Sequence identified in C57Bl/6 obese euglycemic sd7 vs obese sd1

(Identified fragment from 716 to 893 in bold, band size: 178)

5	235	ACCCCTTCTA	CCACTGCCTG	GTTCGAGAAG	TGCCTAAAGA	GCACTGGACC	CCTGAAGGAC
	295	ATAGCATTTGT	TGGGTTTCGCC	ATGTACTATT	TTACCTATGA	CCCATGGATT	GGCAAGTTGC
	355	TGTATCTTGA	AGACTTCTTC	GTGATGAGTG	ATTACAGAGG	CTTTGGTATA	GGATCAGAAA
	415	TTTGAAGAA	TCTAAGCCAG	GTTCCTCATGA	AGTGTGCTG	CAGCAGTATG	CACTTCTTGG
	475	TAGCAGAATG	GAATGAACCA	TCTATCAACT	TCTACAAAAG	AAGAGGTGCT	TCGGATCTGT
10	535	CCAGTGAAGA	GGATGGAGG	CTCTTCAAGA	TGACAAAAGA	GTACTTGCTA	AAAATGGCAG
	595	CAGAGGAGTG	AGGCGTGCCTG	GTGTAGACAA	TGACAACTC	CATTGTGCTT	TAGAATAATT
	655	CTCAGCTTCC	CTTGCTTCT	ATCTTGTGTG	TAGTGAATA	ATAGAGCGAG	CACCCATTCC
	715	<b>AAAGCTTTAT</b>	<b>TACCAGTGAC</b>	<b>GTGTGTGCAT</b>	<b>GTTTGAAATT</b>	<b>CGGTCTGTTT</b>	<b>AAAGTGGCAG</b>
	775	<b>TCATGTATGT</b>	<b>GGTTTGGAGG</b>	<b>CAGAATCTTT</b>	<b>GAACATCTTT</b>	<b>TGATGAAGAA</b>	<b>CAAGGTGGTA</b>
15	835	<b>TGATCTTACT</b>	<b>ATATAAGAAA</b>	<b>AACAAAACCT</b>	<b>CATTCTTGTG</b>	<b>AGTCATTAA</b>	<b>ATGTGTACAA</b>
	895	TGTACACACT	GGTACTTAGA	GTTTCTGTTT	TGATCTTTT	TTTTAAATA	AACTCGCTCT
	955	TTGATTT					

Table 3. Spermidine/spermine N-acetyltransferase Gene Sequence identified in human

20 adipocyte mid-way vs undifferentiated (Identified fragment from 162 to 355 in bold, band size: 149)

	1	CTGGTGTTTA	TCCGTCACTC	GCCGAGGTTT	CTTGGGTCAT	GGTGCCAGCC	TGACTGAGAA
	61	GAGGACGCTC	CCGGGAGACG	AATGAGGAAC	CACCTCCTCC	TACTGTTCAA	GTACAGGGGC
25	121	CTGTCCGCA	AAGGGAAGAA	AAGCAAAAGA	CGAAAATGGC	TAAATTCGTG	ATCCGCCCAG
	181	CCACTGCCGC	CGACTGCAGT	GACATACTGC	GGCTGATCAA	GGAGCTGGCT	AAATATGAAT
	241	ACATGGAAGA	<b>ACAAGTAATC</b>	<b>TTAACTGAAA</b>	<b>AAGATCTGCT</b>	<b>AGAAGATGGT</b>	<b>TTTGGAGAGC</b>
	301	<b>ACCCCTTTTA</b>	<b>CCACTGCCTG</b>	<b>GTTGCAGAA</b>	<b>TGCCGAAAGA</b>	<b>GCACTGGACT</b>	<b>CCGGAAGGTT</b>
	361	ACAGTCTCTA	GCTTCGCCAT	GTACATGGCC	CTTCCGTGTA	CATGGATGGG	CGGGGAGGTA
30	421	ACTAAAAGAT	CCTTTACACA	ATAAAGTAGA	TGATCATGAT	AAATGAGGAC	ACAGCATTTGT
	481	TGGTTTTGCC	ATGTACTATT	TTACCTATGA	CCCGTGGATT	GGCAAGTTAT	TGTATCTTGA
	541	GGACTTCTTC	GTGATGAGTG	ATTATAGAGG	CTTTGGCATA	GGATCAGAAA	TTCTGAAGAA
	601	CTTAAGCCAG	GTGCAATGA	GGTGTGCTG	CAGCAGCATG	CACTTCTTGG	TAGCAGAATG
	661	GAATGAACCA	TCCATCAACT	TCTATAAAAG	AAGAGGTGCT	TCTGATCTGT	CCAGTGAAGA
35	721	GGGTTGGAGA	CTGTTCAAGA	TCGACAGGA	GTACTTGCTA	AAAATGGCAA	CAGAGGAGTG
	781	AGGAGTGCTG	CTGTAGATGA	CAACCTCCAT	TCTATTTTAG	AATAAATTCC	CAACT

Table 4. Human Polyamine Oxidase (CG140122-01) DNA and Protein Sequence

40	CGCGCTCGCCG	CAGACTTACTT	CCCCGGCTCAG	CAGGGAAGGTT	CCTAGAAGGTGAC
	GCGGACGGTAT	GCAGAGTTG	GAATCCAGTGGT	GACAGTGCAGT	GACCCCTCAGTGC
	GGCCTACGGAGA	AGGGGACAGC	CTCGTGGTGGT	GATCGGCGCGGCT	TGGCTGGCCTG
	GCTGCAGCCA	AAGCACTTCT	GTAGCAGGGTT	CAAGGATGCT	CACTGTGCTT
	AGCCACATCG	GAGGCGGTG	GCAGAGTGT	GAAACTTGAC	ACGCCACCTTT
45	GAGCTGGGA	GCCACCTGGAT	CCATGGGAA	CCGCTATCTAT	CTAGCAGAAGCC
	ACGGCCTG	GAAGAGACA	ACCGATGGGAA	CCGAGCGTGGG	CCGCTATCTAT
	AAACCACTG	CAATGCTG	AAAGTCAAA	ATAGCGTGGG	GGTGTTCAC
	CCCGAGAG	GAGGAGG	TGCGT	CAACCGAGAG	GAGGAGG
50	ATCCAGCAGT	ACCTGAAGG	TGGAGAGCT	GTGAGAGCAGT	CACACAGCAT
	TCCCTGAGCG	CTTGGGAGT	TGGACCGAGAT	CCCGCGGCT	CACCATCAT
	GGCTTCATG	CGGCTTGT	GAGCTGCT	GGCGAGGGCAT	CCCTGCCACGT
	GGGAAACCT	GTCCGTG	CATTCACT	TGGGACAGG	CTCAGCCCGCC
55	ATTGAGCCCC	GGGGTGGG	GGCGACCA	ATCAGCACT	GGGAGGGT
	GAGGAGCCCC	GGGGGAGG	TGGGATGAG	GATGAGCAGT	GGTGGTGGT
	GAGGAGCC	TGAGCTGAT	CCCGCGG	ACCATGTGAT	TGACCGT
	GGGAGG	CAGTACAC	CAATTTCT	CCGCGCAGG	CTGCCAGAG
60	ACCTACCA	CACTGAGCT	TGGTACCG	CAAGATCT	GCGGCTT
	GAGCGCT	ACGGCCAT	GTGCTGAG	CGGCTGAT	CGGGGAGG
	AAGTGTGAT	GACGAGG	CAGTGG	CCGAGAT	CTGCA
	AAACCA	CACTTCA	AACTCGG	CAATCTG	CGCTG
	TTCCGTG	GCTCCTAT	TACAC	CGCAGT	GGGCT

GCGAAGCCCTGCGGTACACGGAGAGCTCAAAGACAGCGCCCATGCAGGTGCTGTTTCC  
 GGTGAGGCCACCCACCGAAGTACTATTCCACCACCCACGGTGTCTGCTGTCCGCCAG  
 CGTGAGGCTGCCCGCTCATTGAGATGTACCGAGACCTCTCCAGCAGGGGACCTGAGGG  
 CTGTCTCGTGTCTGAGAAGAGCCACTAACTCGTGACCTCCAGCCTGCCCTTGCTGCCG  
 5 TGTGCTCTGCCTTCTGATCTCTGTAGAAAGGATTTTATCTTCTGTAGAGCTAGCCG  
 CCTGACTGCCTTCAGACCTGGCCCTGTAGCTTT

Table 5. CG140122-01-prot 325 aa

10 MQSCSSGDSADDPLSRGLRRRGQPRVVVIGAGLAGLAAAKALLEQGFDTVTLEASSHI  
 GGRVQSVKLGHATFELGATWIHSGHNPYHLAEANGILLEETDGRSVGRISLYSKNGV  
 ACYLTHNHGRRIPKDVVEEFSDLYNEVYNTQEFFRHKDPVNAESQNSVGVPTRFEEVRNRI  
 RNDPDDPEATKRLKLAAMIQQYLKVESCESSSHSMDEVSLSAFGWEIIPGAHHIIPSGFM  
 15 RVVLLAAGIPAHVILQKGPVRCIHWDQASARPRGPEIEPRGEGDHNHDTGEGGQGGEEP  
 RGRWDEDEQWSVVVECEDRELIPADHVITVSLGVLKRQYTSFFRPGLPTEKVAIHRLL  
 GIGTTDKIFLEFEEPPWGPCNSLQFVWEDEAESHTLTYPPELWYRKICGFDVLYPPERY  
 GHVLSGWI CGBEALVMEKCDDEAVAEICTEMLRQFTGNPNI PKPRRILRSANGSNPYFRG  
 SYSTQVSSGADVEKLAKPLPYTESSTAPMQVLFSGEATHRKYSTTHGALLSGQREA  
 20 ARLIEMRYDLFQOGT

Table 6. Clustal W, Protein Domains, Cellular Location and Locus

The following is an alignment of the protein sequences of CG140122-01 and its  
 alternative spliced variant CG140122-02, which are the equivalent of the public sequences  
 25 AY033889 and BC000669.1, respectively. They are clustalled with the polyamine oxidase  
 of Zea Mays, of which the structural analysis has revealed much of the domain structure of  
 this amine oxidase. The region in bold represents the amine oxidase domain. The dotted  
 region represents the signal peptide.

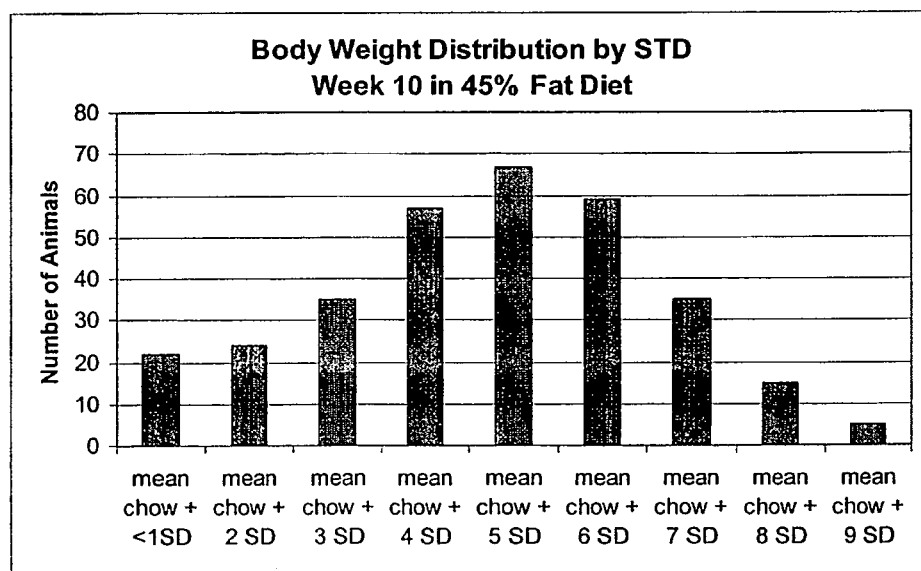
CG140122_01_AY033889	1	MQSCSSGDSADDPLSRGLRRRGQPRVVVIGAGLAGLAAAKALLEQGFDTVT	52
CG140122_02_BC000669.1	1	MQSCSSGDSADDPLSRGLRRRGQPRVVVIGAGLAGLAAAKALLEQGFDTVT	52
Q9FXZ9_PAO_Zea_mays	1	MQSCSSGDSADDPLSRGLRRRGQPRVVVIGAGLAGLAAAKALLEQGFDTVT	60
CG140122_01_AY033889	53	WLEASSHIQGRVQSVKLGHATFELGATWIHSGHNPYHLAEANGILLEETDGRSVGRISLYSKNGV	112
CG140122_02_BC000669.1	53	WLEASSHIQGRVQSVKLGHATFELGATWIHSGHNPYHLAEANGILLEETDGRSVGRISLYSKNGV	112
Q9FXZ9_PAO_Zea_mays	61	WLEASSHIQGRVQSVKLGHATFELGATWIHSGHNPYHLAEANGILLEETDGRSVGRISLYSKNGV	117
CG140122_01_AY033889	113	SLYSKNGVACYLTHNHGRRIPKDVVEEFSDLYNEVYNTQEFFRHKDPVNAESQNSVGVPTRFEEVRNRI	172
CG140122_02_BC000669.1	113	SLYSKNGVACYLTHNHGRRIPKDVVEEFSDLYNEVYNTQEFFRHKDPVNAESQNSVGVPTRFEEVRNRI	172
Q9FXZ9_PAO_Zea_mays	118	SLYSKNGVACYLTHNHGRRIPKDVVEEFSDLYNEVYNTQEFFRHKDPVNAESQNSVGVPTRFEEVRNRI	170
CG140122_01_AY033889	173	RNDPDDPEATKRLKLAAMIQQYLKVESCESSSHSMDEVSLSAFGWEIIPGAHHIIPSGFM	230
CG140122_02_BC000669.1	173	RNDPDDPEATKRLKLAAMIQQYLKVESCESSSHSMDEVSLSAFGWEIIPGAHHIIPSGFM	230
Q9FXZ9_PAO_Zea_mays	171	RNDPDDPEATKRLKLAAMIQQYLKVESCESSSHSMDEVSLSAFGWEIIPGAHHIIPSGFM	230
CG140122_01_AY033889	231	RVVLLAAGIPAHVILQKGPVRCIHWDQASARPRGPEIEPRGEGDHNHDTGEGGQGGEEP	286
CG140122_02_BC000669.1	231	RVVLLAAGIPAHVILQKGPVRCIHWDQASARPRGPEIEPRGEGDHNHDTGEGGQGGEEP	281
Q9FXZ9_PAO_Zea_mays	231	RVVLLAAGIPAHVILQKGPVRCIHWDQASARPRGPEIEPRGEGDHNHDTGEGGQGGEEP	285
CG140122_01_AY033889	287	RGRWDEDEQWSVVVECEDRELIPADHVITVSLGVLKRQYTSFFRPGLPTEKVAIHRLL	346
CG140122_02_BC000669.1	287	RGRWDEDEQWSVVVECEDRELIPADHVITVSLGVLKRQYTSFFRPGLPTEKVAIHRLL	293
Q9FXZ9_PAO_Zea_mays	286	RGRWDEDEQWSVVVECEDRELIPADHVITVSLGVLKRQYTSFFRPGLPTEKVAIHRLL	307
CG140122_01_AY033889	347	GIGTTDKIFLEFEEPPWGPCNSLQFVWEDEAESHTLTYPPELWYRKICGFDVLYPPERY	406
CG140122_02_BC000669.1	347	GIGTTDKIFLEFEEPPWGPCNSLQFVWEDEAESHTLTYPPELWYRKICGFDVLYPPERY	353
Q9FXZ9_PAO_Zea_mays	308	GIGTTDKIFLEFEEPPWGPCNSLQFVWEDEAESHTLTYPPELWYRKICGFDVLYPPERY	358
CG140122_01_AY033889	407	GHVLSGWI CGBEALVMEKCDDEAVAEICTEMLRQFTGNPNI PKPRRILRSANGSNPYFRG	466
CG140122_02_BC000669.1	407	GHVLSGWI CGBEALVMEKCDDEAVAEICTEMLRQFTGNPNI PKPRRILRSANGSNPYFRG	413
Q9FXZ9_PAO_Zea_mays	359	GHVLSGWI CGBEALVMEKCDDEAVAEICTEMLRQFTGNPNI PKPRRILRSANGSNPYFRG	415
CG140122_01_AY033889	467	SYSTQVSSGADVEKLAKPLPYTESSTAPMQVLFSGEATHRKYSTTHGALLSGQREA	526
CG140122_02_BC000669.1	467	SYSTQVSSGADVEKLAKPLPYTESSTAPMQVLFSGEATHRKYSTTHGALLSGQREA	473
Q9FXZ9_PAO_Zea_mays	416	SYSTQVSSGADVEKLAKPLPYTESSTAPMQVLFSGEATHRKYSTTHGALLSGQREA	465
CG140122_01_AY033889	527	ARLIEMRYDLFQOGT	555
CG140122_02_BC000669.1	474	ARLIEMRYDLFQOGT	502
Q9FXZ9_PAO_Zea_mays	465	ARLIEMRYDLFQOGT	500

The variants of the human Polyamine oxidase obtained from direct cloning and/or public databases:

In addition to the human version of the Polyamine oxidase identified as being differentially expressed in the experimental study, no other variants have been identified by  
 5 direct sequencing of cDNAs derived from many different human tissues and from sequences in public databases. The two alternative spliced variants (see clustalW above) are public sequences; no other splice variants have been identified at CuraGen. No SNPs have been found for polyamine oxidase. The preferred variant of all those identified, to be used for screening purposes, is CG140122-01.

10

Figure 1. Bar Graph of Diet induced obesity Under Discovery Process BP24.2.



15

Species #1 mouse      Strains NZB, SM/J, C56Bl/6.  
 Species # 2          Human

20

SPECIES #1. mouse (NZB vs SM/J):

A gene fragment of the mouse spermine/spermidine N-acetyltransferase was initially found to be upregulated by 1.9 fold in the adipose of NZB mice relative to SM/J

mice using CuraGen's GeneCalling™ method of differential gene expression. A differentially expressed mouse gene fragment migrating at approximately 411 nucleotides in length (Figure 1a. - red vertical line) was definitively identified as a component of the mouse spermine/spermidine N-acetyltransferase cDNA in NZB and SM/J mouse strains.

5 The method of competitive PCR was used for conformation of the gene assessment. The chromatographic peaks corresponding to the gene fragment of the mouse spermidine/spermine N-acetyltransferase are ablated when a gene-specific primer (see below) which competes with primers in the linker-adaptors during the PCR amplification. The peaks at 411 nt in length are ablated (green trace) in the sample from both the NZB and

10 the SM/J mice. The altered expression in of these genes in the animal model support the role of Polyamine Oxidase in the pathogenesis of obesity and/or diabetes.

SPECIES #1 mouse (C57Bl/6 obese euglycemic sd7. vs obese sd1):

Figures 3A and 3B show that a differentially expressed gene fragment of the mouse spermine/spermidine N-acetyltransferase was initially found to be upregulated by 1.8 fold

15 in the epididymal fat pad of the obese euglycemic sd7 mice relative to the obese sd1 mice using CuraGen's GeneCalling™ method of differential gene expression. A differentially expressed rat gene fragment migrating at approximately 178 nucleotides in length (Figures 3A and 3B- vertical line) was definitively identified as a component of the mouse

20 spermine/spermidine N-acetyltransferase cDNA in the Troglitazone treated and the untreated SHR control rats (in the graphs, the abscissa is measured in lengths of nucleotides and the ordinate is measured as signal response). The method of competitive PCR was used for conformation of the gene assessment. The electropherogramatic peaks corresponding to the gene fragment of the mouse spermidine/spermine N-acetyltransferase are ablated when

25 a gene-specific primer (see below) which competes with primers in the linker-adaptors during the PCR amplification. The peaks at 178 nt in length are ablated (green trace) in the sample from both the C57Bl/6 obese euglycemic sd7 and obese sd1. mice. The altered expression in of these genes in the animal model support the role of Polyamine Oxidase in the pathogenesis of obesity and/or diabetes.

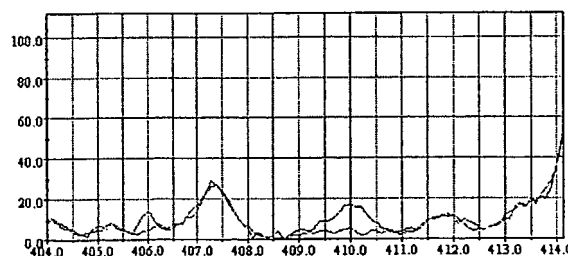
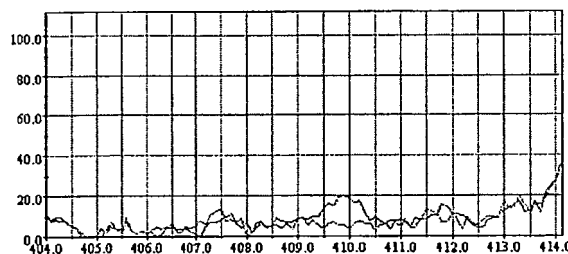
30

SPECIES #2 human (adipocyte mid-way vs undifferentiated):

Figure 4 shows a differentially expressed gene fragment in Discovery Study MB.08 identified in human adipocyte mid-way vs undifferentiated is from the human

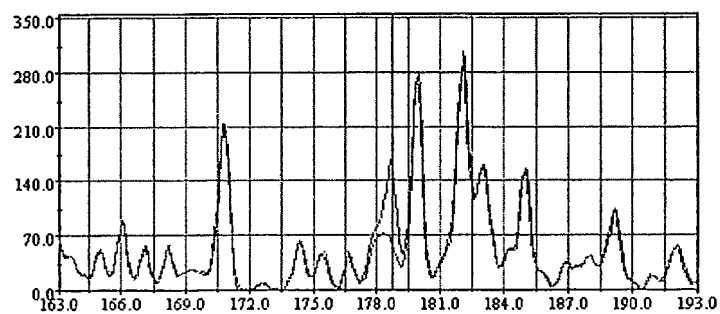
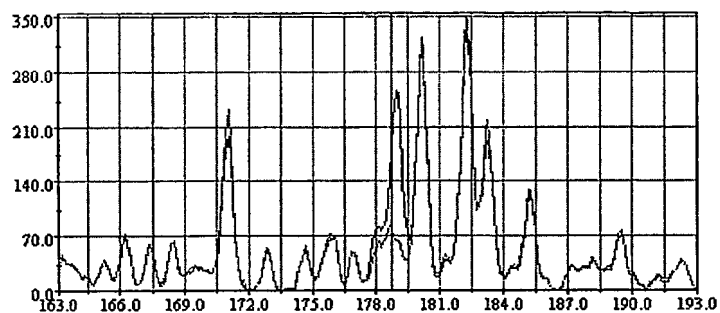
- spermidine/spermine N-acetyltransferase A gene fragment of the human spermine/spermidine N-acetyltransferase was initially found to be upregulated by 1.6 fold in the mid-way human adipocytes relative to the undifferentiated human adipocytes using CuraGen's GeneCalling<sup>TM</sup> method of differential gene expression. A differentially
- 5 expressed human gene fragment migrating at approximately 194 nucleotides in length (Figure 3A - vertical line) was definitively identified as a component of the human spermine/spermidine N-acetyltransferase cDNA in human mid-way differentiated and undifferentiated adipocytes (in the graphs, the abscissa is measured in lengths of
- 10 nucleotides and the ordinate is measured as signal response). The method of competitive PCR was used for conformation of the gene assessment. The chromatographic peaks corresponding to the gene fragment of the human spermine/spermidine N-acetyltransferase are ablated when a gene-specific primer (see below) which competes with primers in the linker-adaptors during the PCR amplification. The peaks at 194 nt in length are ablated
- 15 (green trace) in the sample from both the human mid-way differentiated and undifferentiated adipocytes. The altered expression of these genes in the human cellular model support the role of Polyamine Oxidase in the pathogenesis of obesity and/or diabetes.

- Figures 2A and 2B. Differential Expression of Gene Fragment from Mouse
- 20 Spermidine/spermine N-acetyltransferase.



Figures 3A and 3B. Differentially Expressed Gene Fragment from C57BI/6 Obese Euglycemic sd7.Mouse Spermidine/spermine N-acetyltransferase.

5



10

Figure 4. Differentially Expressed Gene Fragment in Human from Human Spermidine/spermine N-acetyltransferase.

15

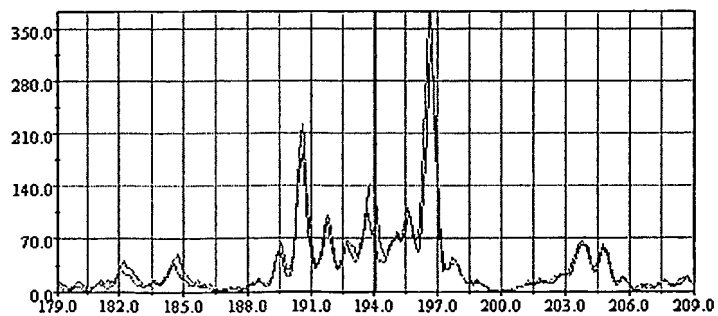
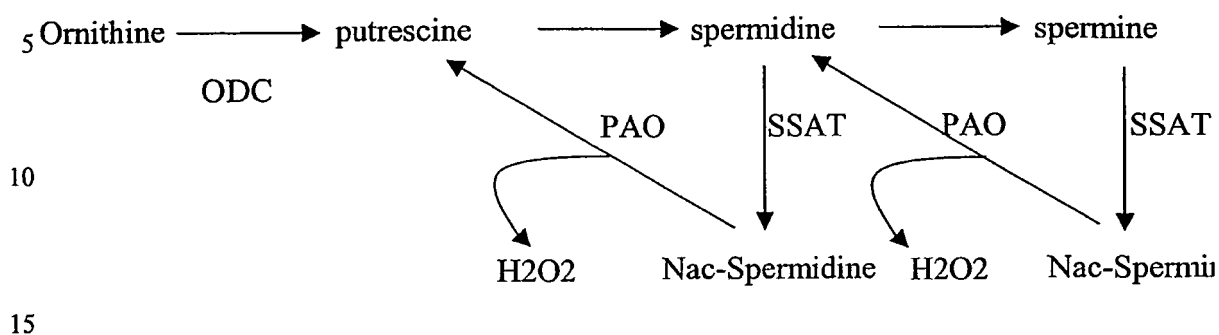


Figure 5. Human Polyamine Oxidase and Assays for Cell Line Expression

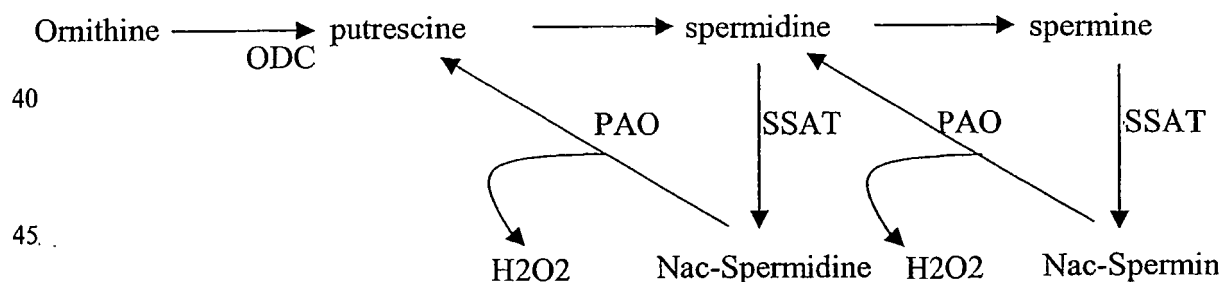


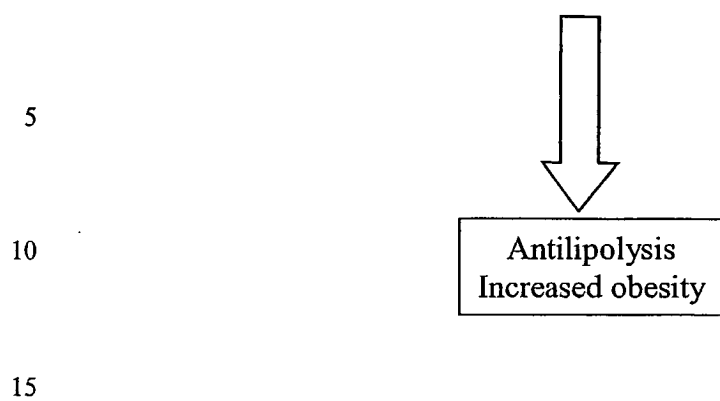
ODC = ornithine decarboxylase  
 PAO = polyamine oxidase  
 SSAT = spermidine/spermine N-acetyltransferase

#### Biochemistry of PAO:

- Catalyses oxidation of secondary amino group of spermine, spermidine and their acetyl derivatives
- Cofactor FAD
- Monomeric

The following illustration suggests how alterations in expression of the human polyamine oxidase and associated gene products function in the etiology and pathogenesis of obesity and/or diabetes. The scheme incorporates the unique findings of these discovery studies in conjunction with what has been reported in the literature. The outcome of inhibiting the action of the human polyamine oxidase would be a way to increase lypolysis by inhibiting anti-lypolytic effects of hydrogen peroxide.





ODC = ornithine decarboxylase  
 PAO = polyamine oxidase  
 20 SSAT = spermidine/spermine N-acetyltransferase

**D. NOV 14a - Human Cytoplasmic Malic Enzyme – CG140316-01**

The present invention discloses novel associations of proteins and polypeptides and the nucleic acids that encode them with various diseases or pathologies. The proteins and related proteins that are similar to them are encoded by a cDNA and/or by genomic DNA.

- 5 The proteins, polypeptides and their cognate nucleic acids were identified by CuraGen Corporation in certain cases. The Cytoplasmic Malic Enzyme -encoded protein and any variants, thereof, are suitable as diagnostic markers, targets for an antibody therapeutic and targets for small molecule drugs. As such the current invention embodies the use of recombinantly expressed and/or endogenously expressed protein in various screens to
- 10 identify such therapeutic antibodies and/or therapeutic small molecules.

**Discovery Process**

- The following sections describe the study design(s) and the techniques used to identify the Cytoplasmic Malic Enzyme - encoded protein and any variants, thereof, as
- 15 being suitable as diagnostic markers, targets for an antibody therapeutic and targets for a small molecule drugs for Obesity and Diabetes.

**Studies:**

- BP24.02 Dietary Induced Obesity in Mice
- 20 MB.04: Genetic Models of Obesity in Mice

**Study Statements:**

- BP24.02: The predominant cause for obesity in clinical populations is excess caloric intake. This so-called diet-induced obesity (DIO) is mimicked in animal models by feeding high
- 25 fat diets of greater than 40% fat content. The DIO study was established to identify the gene expression changes contributing to the development and progression of diet-induced obesity. In addition, the study design seeks to identify the factors that lead to the ability of certain individuals to resist the effects of a high fat diet and thereby prevent obesity. The sample groups for the study had body weights +1 S.D., + 4 S.D. and + 7 S.D. of the chow-
- 30 fed controls (below). In addition, the biochemical profile of the + 7 S.D. mice revealed a further stratification of these animals into mice that retained a normal glycemic profile in spite of obesity and mice that demonstrated hyperglycemia. Tissues examined included hypothalamus, brainstem, liver, retroperitoneal white adipose tissue (WAT), epididymal

WAT, brown adipose tissue (BAT), gastrocnemius muscle (fast twitch skeletal muscle) and soleus muscle (slow twitch skeletal muscle). The differential gene expression profiles for these tissues should reveal genes and pathways that can be used as therapeutic targets for obesity.

5

MB.04: A large number of mouse strains have been identified that differ in body mass and composition. The AKR and NZB strains are obese, the SWR, C57L and C57BL/6 strains are of average weight whereas the SM/J and Cast/Ei strains are lean. Understanding the gene expression differences in the major metabolic tissues from these strains will elucidate the pathophysiological basis for obesity. These specific strains of rat were chosen for differential gene expression analysis because quantitative trait loci (QTL) for body weight and related traits had been reported in published genetic studies. Tissues included whole brain, skeletal muscle, visceral adipose, and liver.

15

Species #1	Mouse Strains	C57BL/6
Species #2	Mouse Strains	NZB, SMJ

Cytoplasmic Malic Enzyme:

20

This gene encodes a cytosolic, NADP-dependent enzyme that generates NADPH for fatty acid biosynthesis. The NADP-dependent malic enzyme (EC 1.1.1.40) has two forms: cytosolic and mitochondrial, that differ significantly in their activity and tissue distribution. The activity of the cytosolic enzyme, the reversible oxidative decarboxylation of malate, links the glycolytic and citric acid cycles. The reaction it catalyzes is:

25

Malate + NADP<sup>+</sup> ⇌ Pyruvate + CO<sub>2</sub> + NADPH

30

Cytoplasmic malic enzyme is one of the anaplerotic reactions, replenishing intermediates of the citrate cycle that are utilized for biosynthesis. It also participates in the pyruvate-citrate shuttle, enabling the export of acetyl CoA from the mitochondrion to cytoplasm for fatty acid synthesis. The regulation of expression for this gene is complex. Increased expression can result from elevated levels of thyroid hormones or by higher proportions of carbohydrates in the diet.

The direct sequence of the nucleotide-long gene fragment and the gene-specific primers used for competitive PCR are indicated on the cDNA sequence of the Cytoplasmic Malic Enzyme and shown below in bold.

### 5 Competitive PCR Primer for the mouse Cytoplasmic Malic Enzyme:

Table 1. Sequence Gene Sequence #1 (fragment from 1520 to 1801 in **bold**. band size: 282)

1039	AAAGGACTAA	TAGTTAAGGG	TCGTGCATCT	CTCACAGAAG	AGAAAGAGGT	GTTCGCCCAT
1099	GAACATGAAG	AAATGAAGAA	TCTGGAAGCC	ATTGTTCAAA	AGATAAAACC	AACATGCCCTC
1159	ATAGGAGTTG	CTGCAATTGG	TGGTGCITTC	ACTGAACAAA	TTCTCAAGGA	TATGGCTGCC
1219	TTCACGAGC	GGCCCATCAT	CTTGTCTTTG	AGTAATCCGA	CCAGCAAAGC	GGAGTGTCTCT
1279	GCAGAGCAGT	GCTACAAGGT	GACCAAGGGA	CGTGCAATCT	TTGCCAGCGG	CAGTCCCTTTT
1339	GATCCAGTCA	CTCTCCCAGA	TGGACGGACT	CTGTTTCCTG	GCCAAAGCAA	CAATTCCCTAC
1399	GTGTTCCCTG	GAGTTGCTCT	TGGGGTGGTG	GCCTGCGGAC	TGAGACACAT	CGATGATAAG
1459	GTCTTCCTCA	CCACTGCTGA	GGTCATATCT	CAGCAAGTGT	CAGATAAACA	CCTGCAAGAA
1519	GGCCGGCTCT	<b>ATCCTCCTTT</b>	<b>GAATACCATT</b>	<b>CGAGGCGTTT</b>	<b>CGTTGAAAAT</b>	<b>TGCAGTAAAG</b>
1579	<b>ATTGTGCAAG</b>	<b>ATGCATACAA</b>	<b>AGAAAAGATG</b>	<b>GCCACTGTTT</b>	<b>ATCCTGAACC</b>	<b>CCAAAACAAA</b>
1639	<b>GAAGAATTG</b>	<b>TCTCCTCCCA</b>	<b>GATGTACAGC</b>	<b>ACTAATTATG</b>	<b>ACCAGATCCT</b>	<b>ACCTGATTGT</b>
1699	<b>TATCCGTGGC</b>	<b>CTGCAGAAGT</b>	<b>CCAGAAAATA</b>	<b>CAGACCAAAG</b>	<b>TCAACCAGTA</b>	<b>ACGCAACAGC</b>
1759	<b>TAGGATTTT</b>	<b>AACTTTATTA</b>	<b>GTAAAATCTT</b>	<b>GAAGTTTCA</b>	<b>TGATCTTTAA</b>	<b>GGGTCAGAAT</b>
1819	CTTTTATGAT	GATTCATAGT	GTGCTTAGAA	TAAGGTGATT	TTAGTTTAAT	AACAACTCA
1879	TGGGAGTCTA	TTAGGATAAA	TTAGGATAAA	TTTCACACCA	GACGGTTTGT	TTTCACTTAC
1939	TGTGGATATT	TATGTTTCT	CTTGTGATTA	TTCTCTTTAT	GAATTCGTGT	TAAAAGCTAC
1999	TGTACCTGCT	GCTGAGAAAG	TCCTCACTGA	TATGTAGGAA	GCTAATGGAA	GACCCACACT
2059	AGTAATAAAT	TAATATAGCA	TAACCTTGATT	ACATTTAATG	CCTACAGTTC	TTTCTTGACT
2119	ATTTTGCTAA	AATCTCTTAA	ACAGAAAAGA	TAAACACAAA	CTTGGGTATA	GCTGAACCTT
2179	TACTAAACAG	AAGCACTACT	TTGTTGCCTA	GAGAAAATCT	TCTCAGGACT	TTTATTCCAG
2239	GCCTCCGTTA	GCTTTGTCT	CTTTGTACAC	CTGACTCAAC	ACC	

30 (gene length is 3105, only region from 1039 to 2281 shown)

Table 2. Sequence #2 Gene Sequence (fragment from 245 to 420 in **bold**. band size: 176)

1	CGCCGGGCGG	CTTGGGGGGC	CGCCGCCCGC	CGGACTCCGC	GTCCGCCCCG	CCACCGGTGC
61	CAGCCATGGA	GCCCGGAGCC	CCCGGCCGCC	GACACACCCA	CCAGCGCGGC	TACCTGCTGA
121	CGCGGGACCC	GCATCTCAAC	AAGGACTTGG	CTTTTACTCT	GGAAGAGAGA	CAGCAGTTGA
181	ACATTCATGG	ATTGTTGCCG	CCCTGCATCA	TCAGCCAGGA	GCTCCAGGTC	CTTAGAATAA
241	TTAAGAAATT	<b>CGAACGACTG</b>	<b>AACTCTGACT</b>	<b>TCGACAGGTA</b>	<b>TCTCCTGTAA</b>	<b>ATGGACCTGC</b>
301	<b>AAGACAGAAA</b>	<b>TGAGAAGCTC</b>	<b>TTCTACAGCG</b>	<b>TGCTCATGTC</b>	<b>TGATGTTGAA</b>	<b>AAGTTCATGC</b>
361	<b>CTATTGTTTA</b>	<b>CACCCCCACC</b>	<b>GTGGGCCTCG</b>	<b>CATGCCAGCA</b>	<b>GTACAGTTTG</b>	<b>GCATTCGGGA</b>
421	AGCCAAGAGG	CCTCTTTATT	AGTATCCATG	ACAAAGGGCA	CATTGCTTCA	GTTCTTAATG
481	CATGGCCAGA	GGATGTCGTC	AAGGCTATTG	TGGTAACTGA	TGGAGAGCGC	ATCCTTGGCT
541	TGGGAGACCT	TGGCTGTAAT	GGGATGGGCA	TCCCTGTGGG	TAAACTGGCC	CTTTACACGG
601	CATGTGGAGG	GGTGAACCCA	CAACAGTGTC	TACCCATCAC	TTTGGATGTG	GGAACAGAAA
661	ATGAGGAGTT	ACTTAAGGAT	CCACTGTACA	TCGGGCTGCG	GCACCGGCGA	GTCAGAGGCC
721	CTGAGTATGA	CGCCTTCTCG	GATGAGTTCA	TGGAGGCAGC	GTCTTCCAAA	TATGGCATGA
781	ATTGCCTTAT	TCAGTTTGAA	GATTTTGCCA	ATCGGAATGC	ATTTGCTCTC	CTGAACAAGT
841	ATCGAAACAA	GTATTGCACA	TTTAACGATG	ATATTCAAGG	AACAGCGTCT	GTTGCGGTTG

(gene length is 3129, only region from 1 to 900 shown)

### 50 Table 3. Human Cytoplasmic Malic Enzyme Gene Sequence.

>CG140316-01 2058 nt  
 ATGGAGCCCGAAGCCCCCGTCCGCCACACCCATCAGCGGGCTACCTGCTGACACGG  
 AACCCCTACCTCAACAAGGACTTGGCCTTTACCCTGGAAGAGAGACAGCAATTGAACATT  
 CATGGATTGTTGCCACCTTCTTCAACAGTCAGGAGATCCAGGTTCTTAGAGTAGTAAAA

AATTTTCGAGCATCTGAACTCTGACTTTGACAGGTATCTTCTCTAATGGATCTCCAAGAT  
 AGAAATGAAAACTCTTTTATAGAGTGTGACATCTGACATTGAGAAATTCATGCCTATT  
 GTTTATATCTCCCACTGTGGGTCTGGCTTGCCAAATATAGTTTGGTGTTCGGAAGCCA  
 AGAGGTCTCTTTATCTATCCAGATCGAGGGCATATTGCTTCAGTTCTCAATGCATGG  
 CCAGAAGATGTCAATCAAGGCCATTGTGGTACTGATGGAGAGCGTATTCTTGGCTTGGGA  
 GACCTTGGCTGTAATGGAATGGGCATCCCTGTGGGTAAATGGCTCTATATACAGCTTGC  
 GGAGGGATGAATCCCAAGAAATGTCTGCCTGTCAATCTGGATGTGGGAACCGAAATGAG  
 GAGTTACTTAAAGATCCACTCTACATTGGACTACGGCAGAGAAGAGTAAGAGGTTCTGAA  
 TATGATGATTTTGGACGAATTCATGGAGGAGTTCTTCCAAGTATGGCATGAATTGC  
 CTTATTCACTTGAAGATTTTGGCAATGTGAATGCATTCGTCTCCTGAACAAGTATCGA  
 AACCAGTATTGCACATTCATGATGATATTCAAGGAACAGCATCTGTGCAAGTTCAGGT  
 CTCCTTGCACTCTTGAATAACCAAGAACAACTGTCTGATCAAACTACTATTCCAA  
 GGAGCTGGAGAGGCTGCCCTAGGGATTGCACACCTGATTGTGATGGCCTTGGAAAAAGAA  
 GGTTTACCAAAAGAGAAAGCCATCAAAAAGATATGGCTGGTTGATTCAAAAGGATTAATA  
 GTTAAGGACGTGCTTCCCTTAACACAGAGAAAGAGAAGTTTGGCCATGAACATGAAGAA  
 ATGAAGAACCTAGAAAGCCATTGTTCAAGAAATAAAACCACTGCCCTCATAGGAGTTGCT  
 GCAATTGGTGGTGCATTCTCAGAACAAATCTCAAAGATATGGCTGCCTTCAATGAACGG  
 CCTATTATTTTGGCTTTGAGTAATCCAAGTACCAAGCAGAAATGTTCTGCAGAGCAGTGC  
 TACAAAATAACCAAGGGACGTGCAATTTTGGCAGTGGCAGTCTTTTGATCCAGTCACT  
 CTTCCAAATGGACAGACCCTATATCTGGCCAGGCAACAATCTACGTGTTCCTGGA  
 GTTGCTCTTGGTGTGTGGCGTGTGGATTGAGGCAGATCAGAGATAATATTTCTCTCACT  
 ACTCTGAGGTATAGCTCAGCAAGTGTGAGATAAACTTGGAGAGGGTGGGCTTTAT  
 CCTCCTTTGAATACCAATAGAGATGTTCTCTGAAAATTCAGAAAAGATTGTGAAAAGAT  
 GCATACCAAGAAAAGACAGCCACAGTTTATCTGAACCGCAAAACAAAGAACATTTGTC  
 CGCTCCCAGATGTATAGTACTGATTATGACCAGATTCTACCTGATTGTTATTCTTGGCCT  
 GAAGAGGTGACAGAAAATACAGACCAAGTTGACCAGTAGGATAATAGCAACATTTCTAA  
 CTCATTATATGAGTCTTTAAACCTTTTATAATTTTAAAGGTGGAAATCTTTTATAATG  
 ATTCTAAGACACTTAGATTAAAGATTTTACTTTAACAGTCTAAAAATTGATAGAGAATA  
 TCGATATAAATTTGGGATAAAACATCAGATGAGACAAATTTGCTTCACTTTGCTTCTGGTT  
 ATTTATGTTTCTGTCTGAATTTCTGCCTACGTTCTCTTAAAGCTGTGTGACGTAC  
 TACGGAGAACTCATCATTTTATACAGGACACTAATGGGAAGACCAAAATTACTAATAA  
 ATTGAATAACCAACATT

Table 4. Amino acid sequence of Human Cytoplasmic Malic Enzyme Protein Sequence

ORF Start: 1 ORF Stop: 1717 Frame: 1

## Human Cytoplasmic Malic Enzyme Protein Sequence:

5 >CG140316-01-prot 572 aa  
 MEPEAPRRRHHQRYLLTRNPHLNKDLAFTLEERQQLNIHGLPPSFNSQEIQVLRVVK  
 NFEHLNSDFDRYLLMLDQDRNEKLFYRVLTSDEKFMPIVYTPVGLACQYSLVFRKP  
 10 RGLFTIHDRGHIASVLNAPEDVIKAIVVTDGERILGLDGLCNGMGI PVGKLALYTAC  
 GGMNPQECPLPVILDVGTENEELLKDPLYIGLRQRRVRGSEYDDFLDEFMEAVSSKYGMNC  
 LIQPEDFANVNAFRLNLYRNQYCTFNDDIQGTASVAVAGLLAALRITKNKLSQQTILFQ  
 GAGEAALGIAHLIVMALEKEGLPKKAIKKIWLVDKGLIVKGRASLTQEKKEFAHEHEE  
 MKNLEAIVQEI KPTALIGVAAIGGAFSEQILKDMAAFNERPIIFALSNPTS KAEC SAEQC  
 15 YKITKGRAIFASGSPFDPVTL PNGQTL YFGQNN SYVFP GVALGVVACGLRQITDNI FLT  
 TAEVIAQQVSDKHLBEGRLYPPLNTIRDVSLKIAEKIVKDAYQEXTATVYPEPQNEAFV  
 RSQMYSYDQILPDCYSWPPEVQKIQTQVDQ

Table 5. Clustal W, Protein Domains, Cellular Location and Locus

The following is an alignment of the protein sequences of the human (CG140316-01), mouse (BC011081.1) and pig (X93016.1) versions of the Cytoplasmic Malic Enzyme. Also included are a variant of this enzyme cloned from liver (CG140316-02) and the mitochondrial NADP-dependent malic enzyme (X79440.1). The domain delineated by the bold line indicates the malic enzyme domain.

25

	CG140316-02	1	-----MEFSAFRRRHTHQVYLLTRPHILN	25
	CG140316-01	1	-----MEFSAFRRRHTHQVYLLTRPHILN	25
	BC011081.1	1	-----MEFSAFRRRHTHQVYLLTRPHILN	25
	X93016.1	1	-----MEFSAFRRRHTHQVYLLTRPHILN	25
	X79440.1	1	MGAALGTGTRLAPWGRACGA LPRWTTTAPAQGHSGKGPAPVPVPLKKSYDYSTRPHILN	40
5	CG140316-02	26	HDLAFTLEERQQLNIHQLPPSPNDCSEVVLWACTHFEHLNSDFEYLLHQLQDPNEIL	45
	CG140316-01	26	HDLAFTLEERQQLNIHQLPPSPNDCSEVVLWACTHFEHLNSDFEYLLHQLQDPNEIL	45
	BC011081.1	26	HDLAFTLEERQQLNIHQLPPSPNDCSEVVLWACTHFEHLNSDFEYLLHQLQDPNEIL	45
	X93016.1	1	-----HDLAFTLEERQQLNIHQLPPSPNDCSEVVLWACTHFEHLNSDFEYLLHQLQDPNEIL	10
	X79440.1	61	HDLAFTLEERQQLNIHQLPPSPNDCSEVVLWACTHFEHLNSDFEYLLHQLQDPNEIL	120
	CG140316-02	86	FYRVLTSQIEEFNFIVTPTVOLACQVYCLVFAIFRRLFITIHGQHIASVLAHAFEEVLI	145
	CG140316-01	86	FYRVLTSQIEEFNFIVTPTVOLACQVYCLVFAIFRRLFITIHGQHIASVLAHAFEEVLI	145
	BC011081.1	86	FYRVLTSQIEEFNFIVTPTVOLACQVYCLVFAIFRRLFITIHGQHIASVLAHAFEEVLI	145
	X93016.1	11	-----FYRVLTSQIEEFNFIVTPTVOLACQVYCLVFAIFRRLFITIHGQHIASVLAHAFEEVLI	70
	X79440.1	121	FYRVLTSQIEEFNFIVTPTVOLACQVYCLVFAIFRRLFITIHGQHIASVLAHAFEEVLI	180
10	CG140316-02	146	KAIWVTEDEPIQLQDLOCNQHGI PPGCLALYTACQGVNHPQCLPVILQVOTENEELLIC	205
	CG140316-01	146	KAIWVTEDEPIQLQDLOCNQHGI PPGCLALYTACQGVNHPQCLPVILQVOTENEELLIC	205
	BC011081.1	146	KAIWVTEDEPIQLQDLOCNQHGI PPGCLALYTACQGVNHPQCLPVILQVOTENEELLIC	205
	X93016.1	71	-----KAIWVTEDEPIQLQDLOCNQHGI PPGCLALYTACQGVNHPQCLPVILQVOTENEELLIC	130
	X79440.1	181	KAIWVTEDEPIQLQDLOCNQHGI PPGCLALYTACQGVNHPQCLPVILQVOTENEELLIC	240
	CG140316-02	206	PLYIOLQRRPVQS EYLCFLDFEPMBAVSSINQNMNCLIQFEQFANQIAPFLINAYTRISCT	265
	CG140316-01	206	PLYIOLQRRPVQS EYLCFLDFEPMBAVSSINQNMNCLIQFEQFANQIAPFLINAYTRISCT	265
	BC011081.1	206	PLYIOLQRRPVQS EYLCFLDFEPMBAVSSINQNMNCLIQFEQFANQIAPFLINAYTRISCT	265
	X93016.1	131	-----PLYIOLQRRPVQS EYLCFLDFEPMBAVSSINQNMNCLIQFEQFANQIAPFLINAYTRISCT	190
	X79440.1	241	PLYIOLQRRPVQS EYLCFLDFEPMBAVSSINQNMNCLIQFEQFANQIAPFLINAYTRISCT	300
	CG140316-02	266	FNDDIQQTASVAVAGLLAALRITENHLSDTILFQQAQEAALQIAHLIQAHEEQLRKE	325
	CG140316-01	266	FNDDIQQTASVAVAGLLAALRITENHLSDTILFQQAQEAALQIAHLIQAHEEQLRKE	325
	BC011081.1	266	FNDDIQQTASVAVAGLLAALRITENHLSDTILFQQAQEAALQIAHLIQAHEEQLRKE	325
	X93016.1	191	-----FNDDIQQTASVAVAGLLAALRITENHLSDTILFQQAQEAALQIAHLIQAHEEQLRKE	250
	X79440.1	301	FNDDIQQTASVAVAGLLAALRITENHLSDTILFQQAQEAALQIAHLIQAHEEQLRKE	360
20	CG140316-02	326	KATIKIWLVDSDQILVDPASLTQEREKFAHEHEEMHNLEAIVQIIPPTALIQVAAIQA	385
	CG140316-01	326	KATIKIWLVDSDQILVDPASLTQEREKFAHEHEEMHNLEAIVQIIPPTALIQVAAIQA	385
	BC011081.1	326	KATIKIWLVDSDQILVDPASLTQEREKFAHEHEEMHNLEAIVQIIPPTALIQVAAIQA	385
	X93016.1	251	-----KATIKIWLVDSDQILVDPASLTQEREKFAHEHEEMHNLEAIVQIIPPTALIQVAAIQA	310
	X79440.1	361	KATIKIWLVDSDQILVDPASLTQEREKFAHEHEEMHNLEAIVQIIPPTALIQVAAIQA	420
	CG140316-02	386	FEEQLRGMAAFHEPPIIFALDIPTTCAECSAEQVYVITDPAIFASQSPFQVTLRQ	445
	CG140316-01	386	FEEQLRGMAAFHEPPIIFALDIPTTCAECSAEQVYVITDPAIFASQSPFQVTLRQ	445
	BC011081.1	386	FEEQLRGMAAFHEPPIIFALDIPTTCAECSAEQVYVITDPAIFASQSPFQVTLRQ	445
	X93016.1	311	-----FEEQLRGMAAFHEPPIIFALDIPTTCAECSAEQVYVITDPAIFASQSPFQVTLRQ	370
	X79440.1	421	FEEQLRGMAAFHEPPIIFALDIPTTCAECSAEQVYVITDPAIFASQSPFQVTLRQ	480
25	CG140316-02	446	TLFPGQNNHYVFPVALQVWAGQLHIDQVFLITTAEVISQVSDHQBQVLYPPIINT	505
	CG140316-01	446	TLFPGQNNHYVFPVALQVWAGQLHIDQVFLITTAEVISQVSDHQBQVLYPPIINT	505
	BC011081.1	446	TLFPGQNNHYVFPVALQVWAGQLHIDQVFLITTAEVISQVSDHQBQVLYPPIINT	505
	X93016.1	371	-----TLFPGQNNHYVFPVALQVWAGQLHIDQVFLITTAEVISQVSDHQBQVLYPPIINT	430
	X79440.1	481	TLFPGQNNHYVFPVALQVWAGQLHIDQVFLITTAEVISQVSDHQBQVLYPPIINT	540
30	CG140316-02	506	IPDVCLIAEMLVQAYCEITATVYEPQNEEAFVPSQMYSTNTQILICQVWFAEBC	565
	CG140316-01	506	IPDVCLIAEMLVQAYCEITATVYEPQNEEAFVPSQMYSTNTQILICQVWFAEBC	565
	BC011081.1	506	IPDVCLIAEMLVQAYCEITATVYEPQNEEAFVPSQMYSTNTQILICQVWFAEBC	565
	X93016.1	431	-----IPDVCLIAEMLVQAYCEITATVYEPQNEEAFVPSQMYSTNTQILICQVWFAEBC	490
	X79440.1	541	IPDVCLIAEMLVQAYCEITATVYEPQNEEAFVPSQMYSTNTQILICQVWFAEBC	600
	CG140316-02	566	IQTVFHS	572
	CG140316-01	566	IQTVFHS	572
	BC011081.1	566	IQTVFHS	572
	X93016.1	491	-----IQTVFHS	496
	X79440.1	601	IQTVFHS	604

## Human Cytoplasmic Malic Enzyme:

40 572 aa

Locus: 6q12 (syntenic to mouse quantitative trait locus correlated with percentage of body fat. Ref: Mehrabian et al., J Clin Invest 1998; 101(11): 2485-2496)

## Intracellular

45 In addition to the human version of the Cytoplasmic Malic Enzyme identified as being differentially expressed in the experimental study, one other variant has been identified by direct sequencing of cDNAs derived from many different human tissues and from sequences in public databases (CG140316-02, Figure 1C). No splice-form variants have been identified at CuraGen nor were any SNPs identified. The preferred variant of all  
50 those identified, to be used for screening purposes, is CG140316-01.

**Biochemistry and Cell Line Expression:**

The following illustrations summarizes the biochemistry surrounding the human Cytoplasmic Malic Enzyme and potential assays that may be used to screen for antibody therapeutics or small molecule drugs to treat obesity and/or diabetes. Generation of the  
 5 reducing equivalents in form of NADPH may be coupled to enzymatic or fluorescent detection systems to provide a readout of the screening.



Cell lines that express the Cytoplasmic Malic Enzyme include PC-3, CaCo-2 and A549, as  
 10 seen in the RTQ-PCR results shown in Table 6. These and other Cytoplasmic Malic Enzyme expressing cell lines could be used for screening purposes.

**Findings:**

15 The following is a summary of the findings from the discovery studies, supplementary investigations and assays that also incorporates knowledge in the scientific literature. Taken in total, the data indicates that an inhibitor/antagonist of the human Cytoplasmic Malic Enzyme would be beneficial in the treatment of obesity and/or diabetes.

- 20
1. Cytoplasmic malic enzyme is upregulated in both liver and adipose of obese mice in different studies.
  2. Upregulation of cytoplasmic malic enzyme promotes fatty acid synthesis and anaplerotic reactions replenishing TCA cycle.
  - 25 3. Inhibiting cytoplasmic malic enzyme will decrease lipid synthesis and force utilization of stored fatty acids for energy generation.
  4. An inhibitor of this enzyme would therefore be an effective therapeutic for obesity.

**SPECIES #1 (ngsd7 vs. sd1 liver):**

30 Figures 1A and 1B show that a gene fragment of the mouse Cytoplasmic Malic Enzyme was initially found to be up-regulated by 4 fold in the liver tissues of obese mice fed a high fat diet relative to mice resistant to weight gain (on the same diet) using CuraGen's GeneCalling® method of differential gene expression. A differentially

expressed mouse gene fragment migrating, at approximately 283 nucleotides in length (Figure 1A. - vertical line) was definitively identified as a component of the mouse Cytoplasmic Malic Enzyme cDNA (in the graphs, the abscissa is measured in lengths of nucleotides and the ordinate is measured as signal response). The method of competitive

5. PCR was used for conformation of the gene assessment. The electropherogramatic peaks corresponding to the gene fragment of the mouse Cytoplasmic Malic Enzyme are ablated when a gene-specific primer (see below) competes with primers in the linker-adaptors during the PCR amplification. The peaks at 283 nt in length are ablated (green trace) in the sample from both the obese and non-obese mice.

10

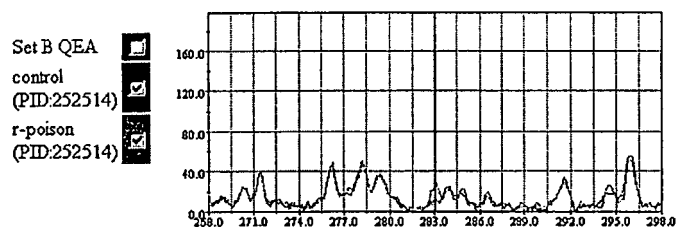
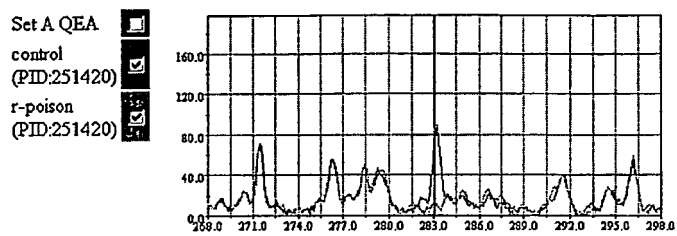
SPECIES #2 (NZB vs. SMJ adipose):

Figures 2A and 2B show that a gene fragment of the mouse Cytoplasmic Malic Enzyme was also found to be up-regulated by 3.2 fold in the adipose of obese NZB mice relative to lean SMJ mice using CuraGen's GeneCalling® method of differential gene expression. A differentially expressed mouse gene fragment migrating, at approximately 175.9 nucleotides in length (Figure 2A. - vertical line) was definitively identified as a component of the mouse Cytoplasmic Malic Enzyme cDNA (in the graphs, the abscissa is measured in lengths of nucleotides and the ordinate is measured as signal response). The method of competitive PCR was used for conformation of the gene assessment. The electropherogramatic peaks corresponding to the gene fragment of the mouse Cytoplasmic Malic Enzyme are ablated when a gene-specific primer (see below) competes with primers in the linker-adaptors during the PCR amplification. The peaks at 175.9 nt in length are ablated (green trace) in the sample from both the obese and non-obese mice.

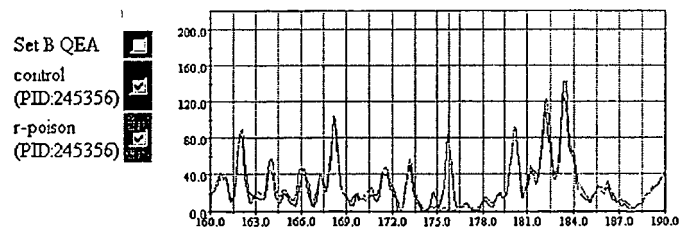
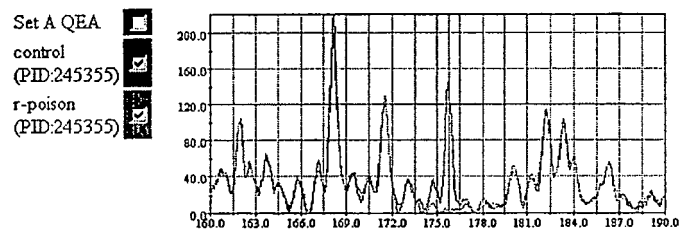
25

Figures 1A and 1B. . Sequence #1. Differentially Expressed Mouse Cytoplasmic Malic Enzyme Gene Fragment.

30



5 Figures 2A and 2B. Sequence #2. Differentially Expressed Mouse Cytoplasmic Malic Enzyme Gene Fragment.



**E. NOV15a - Human ATP Citrate Lyase – CG142427-01, CG142427-02, CG142427-03 and CG142427-04**

The present invention discloses novel associations of proteins and polypeptides and the nucleic acids that encode them with various diseases or pathologies. The proteins and related proteins that are similar to them are encoded by a cDNA and/or by genomic DNA. The proteins, polypeptides and their cognate nucleic acids were identified by CuraGen Corporation in certain cases. The ATP Citrate Lyase-encoded protein and any variants, thereof, are suitable as diagnostic markers, targets for an antibody therapeutic and targets for small molecule drugs. As such the current invention embodies the use of recombinantly expressed and/or endogenously expressed protein in various screens to identify such therapeutic antibodies and/or therapeutic small molecules.

**Discovery Process**

The following sections describe the study design(s) and the techniques used to identify the ATP Citrate Lyase - encoded protein and any variants, thereof, as being suitable as diagnostic markers, targets for an antibody therapeutic and targets for a small molecule drugs for obesity and/or diabetes.

**Studies:**

MB.04: Lean vs. Obese Genetic mouse model

**Study Statements:**

MB.04 : A large number of mouse strains have been identified that differ in body mass and composition. The AKR and NZB strains are obese, the SWR, C57L and C57BL/6 strains are of average weight whereas the SM/J and Cast/Ei strains are lean. Understanding the gene expression differences in the major metabolic tissues from these strains will elucidate the pathophysiologic basis for obesity. These specific strains of rat were chosen for differential gene expression analysis because quantitative trait loci (QTL) for body weight and related traits had been reported in published genetic studies. Tissues included whole brain, skeletal muscle, visceral adipose, and liver.

30

Species #1: Mouse Strains NZB vs SMJ, C57L, Cast, SWR

### ATP Citrate Lyase:

ATP citrate-lyase is the primary enzyme responsible for the synthesis of cytosolic acetyl-CoA in many tissues. has a central role in de novo lipid synthesis. in nervous tissue it may be involved in the biosynthesis of acetylcholine. Figure 1 shows a differentially  
5 expressed gene fragment from the mouse ATP Citrate Lyase.

### Competitive PCR Primer for the Human ATP Citrate Lyase

Confirmatory Result – Human ATP Citrate Lyase (Discovery Study MB.04):

10

Table 1. Human ATP Citrate Lyase Gene Sequence  
(Identified fragment from 1213 to 1277 in *italic*. band size: 65)

1 CTGGGTTGTTTATCGATTTTACTCGATGGCCGATGCCCATGATCAGCTTCCCCTCCTTCTTCATCTTGTGACGAACTCC  
15 81 ATGGGAATGATGCCGCTGTCAAAGGCTTTACTGAACATCTTTGCTGCGGCATCCAAGGCACCCCAACCGGTCTCCAAT  
161 GGTGAGCAGCCCTGAGGTGAGGCTGGAGACCAGGTCCCTTCCAGCCCCAGCACAGATGATGGTGTATGGGCTCCAGAGA  
241 CAGCTGGCCCCGTGATCAGCTGTGACCATCAGACACATCTCAATGAACTGGCAGGAATACTTGGGCAACCTTCTCTGGAAC  
321 CAGAGGAGGCCGAGGACACCCGATGCCATCTCCTCCTTGAAGACCTCGGTGATGGGCATGCCGCATAAATGAGCTC  
401 CTGCCCTCGCTCATCAAGATGCTGGTCAATGAGGCAGGTTTTCGGATCAAACCCAGCTCTCTGGCCCCAAGAGTAGT  
20 481 CCATGGGCACTGTTGGAGGTGGCACTTCTTGGGCGAGTACAATGGCTCCTTTGGCCACCAGATCTTCATACACAGACTGA  
561 ATGATTTCTCCAAGCTCATCGAAGCTTCGGGGCACAAACACTCCTGCTTCTTCAAGGCCTGGTCTTGGCTACTGCACT  
641 TTCAGAAGTCTGGTTGGCAAGCTCCAGCATGGCCAAACTGGACCTCGGAGGAGAACATGGTGGCACAGGTCCCGATAC  
721 ACCAGCAGACCACTGGCTTGGTGGGCGGCCCTCCTTGATGCCCGGCGAGATCTTATATTCTCTGTGCCCTTATCTCC  
801 CCAAGAACTACGATCATCTTGACTCTGGAGTGTCTGGTAGCGCAGCACTGATCCATGAATGTGGACCCAGGGTACCT  
25 881 GTCCCCGCCGATGGCCACGCCCTCATAGACCACTCTGTGGTCCGGGAGATGATGTTATTGAGTTCATTAGACATGCCCTC  
961 CTGAACGTGAGACGTAGGCCACCGTGCCTGGGCGGTACAGTTTGGAGGCCAGGATGTTGTCCAGCA

Table 2. Nucleotide and protein sequence of Human ATP Citrate Lyase

30

CG142427-01

GGCAGAGGCCGGGACAAAAGCCGATCCCGGGAAGCTACCGGCTGCTGGGGTGCTCCGGATTTTGCGGG  
GTTCTGTCGGGCTGTGGAAGAAGCGCCGCGCACGGACTTCGGCAGAGGTAGAGCAGGTCTCTCTGCAGCC  
35 ATGTCGGCCAAGGCAATTTCAAGTATGCTCGGGTCACTCCTGACACAGACTGGGCCCCGCTTGTGCAGGACCA  
CCATCCAGAATCGGTTCAAGTATGCTCGGGTCACTCCTGACACAGACTGGGCCCCGCTTGTGCAGGACCA  
CCCCTGGCTGCTCAGCCAGAACTTGGTAGTCAAGCCAGACCAGCTGATCAAACGTCGTGGAAAACTTGGT  
CTCGTTGGGGTCAACCTCACTCTGGATGGGGTCAAGTCTGGCTGAAGCCACGGCTGGGACAGGAAGCCA  
40 CAGTTGGCAAGGCCACAGGCTTCCTCAAGAACTTTCTGATCGAGCCCTTCGTCCCCCAGCTCAGGCTGA  
GGAGTTCATGTCTGCATCTATGCCACCCGAGAAGGGGACTACGTCCTGTTCCACCACGAGGGGGGTGTG  
GACGTGGTGATGTGGACGCCAAGGCCAGAAAGCTGCTTGTGGCGTGGATGAGAACTGAATCCTGAGG  
ACATCAAAAAACCTGTTGGTCCACGCCCTGAAGACAAGAAAGAAATCTGGCCAGTTTATCTCCGG  
CCTCTCAATTTCTACGAGGACTTGACTTTCACCTACCTCGAGATCAATCCCCTTGATGTGACCAAGAT  
GGAGTCTATGTCCTTGACTTGGCGGCCAAGGTGGACGCCACTGCCGACTACATCTGCAAGTGAAGTGGG  
45 GTGACATCGAGTTCCTCCCCCTTCGGGCGGGAGGCATATCCAGAGGAAGCCTACATTGCAGACCTCGA  
TGCCAAAAGTGGGGCAAGCCTGAAGCTGACCTTGTGAACCCCAAGGGAGGATGTGGACCATGGTGGCC  
GGGGTGGCGCCTCTGTCTGTACAGCGATACCATCTGTGATCTAGGGGGTGTCAACGAGCTGGCAAACT  
ATGGGGAGTACTCAGGCGCCCCAGCGAGCAGCAGACCTATGACTATGCCAAGACTATCCTCTCCCTCAT  
GACCCGAGAGAAGCAACCAAGATGCTCATCATTTGGAGGCGAGCATCGCAAACTTCACCAACGCTG  
50 GCTGCCAGTTCAAGGGCATCGTGAGAGCAATTGAGATTACCAAGGGCCCCCTGAAGGAGCACGAAGTCA  
CAATCTTTGTCCGAAGAGGTGGCCCCAATATCAGGAGGGCTTACGGGTGATGGGAGAAGTCCGGAAGAC  
CACTGGATCCCCATTCATGTCTTGGCACAGAGACTCATATGACGGCCATTGTGGGCATGGCCCTGGGC  
CACCGGCCATCCCCAACAGCCAGCCACACAGCGGCCCACTGCAAACTTCCTCTCAACGCCAGCGGGA  
GCACATCGACGCCAGCCCCAGCAGGACAGCATCTTTTCTGAGTCCAGGGCCGATGAGGTGGCGCTGC  
55 AAAGAAGGCCAAGCCTGCCATGCCACAAGATTCACTCCCAAGTCCAAGATCCCTGCAAGGAAGAGCACC  
ACCTCTTCAGCCCGCACCAAGGCCATTGTGTGGGGCATGCAGACCCGGGCCGTGCAAGGCATGCTGG  
ACTTTGACTATGTCTGCTCCCGAGACGACCCCTCAGTGGCTGCCATGGTCTACCCCTTCACTGGGACCA

CAAGCAGAAGTTTACTGGGGGCACAAAGAGATCCTGATCCCTGTCTTCAAGAACATGGCTGATGCCATG  
 AGGAAGCATCCGGAGGTAGATGTGCTCATCAACTTTGCCTCTCTCCGCTCTGCCTATGACAGCACCATGG  
 AGACCATGAACATATGCCAGATCCGGACCATCGCCATCATAGCTGAAGGCATCCCTGAGGCCCTCACGAG  
 AAAGCTGATCAAGAAGGCGGACCAAGAGGAGTGACCATCATCGGACCTGCCACTGTTGGAGGCATCAAG  
 5 CCTGGGTGCTTTAAGATTGGCAACACAGGTGGGATGCTGGACAACATCCTGGCCTCCAAACTGTACCGCC  
 CAGGCAGCGTGGCCTATGTCTCAGTTCCGGAGGCATGTCCAAGAGCTCAACAATATCATCTCTCGGAC  
 CACGGATGGCGTCTATGAGGGCGTGGCCATTGGTGGGGACAGGTACCCGGGCTCCACATTATGGATCAT  
 GTGTTACGCTATCAGGACACTCCAGGAGTCAAAATGATTGTGGTTCTTGGAGAGATTGGGGCACTGAGG  
 10 AATATAAGATTTCGCCGGGCATCAAGGAGGCGCGCTCACTAAGCCCATCGTCTGTGTTGATCGGGAC  
 GTGTGCCACCATGTTCTCTCTGAGGTCCAGTTTGGCCATGCTGGAGCTTGTGCCAACAGGCTTCTGAA  
 ACTGCAGTAGCCAAAGAACAGGCTTTGAAGGAAGCAGGAGTGTGTTGCCCCGGAGCTTTGATGAGCTTG  
 GAGAGATCATCCAGTCTGTATACGAAGATCTCGTGGCCAATGGAGTCATTGTACCTGCCCAGGAGGTGCC  
 15 GCCCCCAACCGTGCCCATGGACTACTCTGGGCCAGGGAGCTTGGTTTGTATCCGCAACCTGCCTCGTTC  
 ATGACCAGCATCTGCGATGAGCGAGGACAGGAGCTCATCTACGCGGGCATGCCATCACTGAGGTCTTCA  
 AGGAAGAGATGGGCATTGGCGGGGTCTCGGCCCTCTGTTTCCAGAAAAGGTTGCCTAAGTACTCTTG  
 CCAGTTTCATTGAGATGTGTCTGATGGTGACAGCTGATCACGGGCCAGCCGTCTCTGGAGCCCAACACC  
 ATCATTTGTGCGCGAGCTGGGAAAGACCTGGTCTCCAGCCTCACCTCGGGGCTGCTCACCATCGGGGATC  
 20 GGTTTGGGGGTGCTTGGATGCAGCAGCCAAGATGTTCAAGTAAAGCCTTTGACAGTGGCATTATCCCCAT  
 GGAGTTTGTGAACAAGATGAAGAAGGAAGCTGATCATGGGCATTGGTCAACGAGTGAAGTCGATA  
 AACAAACCAGACATGCGAGTGCAGATCTCTAAAGATTACGTACAGGCAGCACTTCCCTGCCACTCTCTGTC  
 TCGATTATGCACTGGAAGTAGAGAAGATTACCACCTCGAAGAAGCCAAATCTTATCCTGAATGTAGATGG  
 TCTCATCCGAGTCGCATTGTGTAGCATGCTTAGAAACTGTGGTCTTCTTACTCGGGAGGAAGCTGATGAA  
 25 TATATTGACATTGGAGCCCTCAATGGCATCTTTGTGCTGGGAAGGAGTATGGGGTTTATTGGACACTATC  
 TTGATCAGAAGAGGCTGAAGCAGGGGCTGTATCGTTCATCCGTGGGATGATATTTATATGTTCTTCCGGA  
 ACACATGAGCATGTAAACAGAGCCAGGAACCCCTACTGCAGTAACTGAAGACAAGATCTCTCCCCCAAGA  
 AAAAGGTACAGACAGCTGGCAGTGGAGCCTGCTTTATTAGCAGGGGCTGGAATGTAAACAGCCACTG  
 GGGTACAGCCACCGAAGACCAACATCCACAGGCTAACACCCCTTCAGTCCACACAAAGAAGCTTCATATT  
 30 TTTTATTAAAGCATAGAAATAAAACCAAGCCAATATTGTGACTTGTCTGCTACCTGCTGATTTTAT  
 TATATGGAAGCATCTAAGTACTGTGAGGATGGGGTCTTCTCATTGTAGGGGCTTAGGATGTGCTTTCT  
 TTTTCCATTAGTTAAACATTTTCTCTCTTTGGAGGAAGGGAATGAAACATTTATGGCCTCAAGATACT  
 ATACATTTAAAGCACCCCAATGTCTCTTTTTTTTTTTTACTTCCCTTTCTTCTCTTATATAACAT  
 35 GAGAACATTGTATTATCTGATTTTAAAGATCTTTTGTATGTACGTGTTAAGGGCTTGTGTTGTTAT  
 CCCACTGAAATGTTCTGTGTTGCAGACCAGAGTCTGTTTATGTCAGGGGGATGGGGCCATTGCATCCTTA  
 GCCATTGTACAAAAATATGTGGAGTAGTAACCTAATATGTAAGTTGTAACATACATACATTTAAATGG  
 AAATGCAGAAAGCTGTGAAATGCTTGTGCTTATGTTCTCTGTATTTATGCAGCTGATTGTCTGTCTG  
 40 TAACTGAAGTGTGGGTCCAAGGACTCCTAATCTTGTGATCTGTAATCCACAAAGATTCTGGGCAGCTG  
 CCACCTCAGTCTCTCTCTGATTATCATAGTCTGGTTTAAATAAACTATATAGTAACAAAAA  
 AA  
 AAAAAAAAAAAAAAAAAA

Table 3. Amino acid sequence of Human ATP Citrate Lyase

ORF Start: 141 ORF Stop: 3444 Frame: 3

45 Human ATP Citrate Lyase Protein Sequence:

**CG142427-01-prot** 1101 aa

MSAKAISEQTGKELLYKFICTTSAIQNRFKYARVTPDWDWARLLQDHPWLLSQNLVVKPD  
 QLIKRRGKLGVLGVNLTLDGVKSWLKPRLGQEATVGKATGFLKNFLIEFPVPHSQAEFFYV  
 50 CIYATREGDYLFLHHEGGVDVGDVDAKAQKLLVGVEKLNPEDIKKHLLVHAPEDKKEIL  
 ASFISGLNFYEDLYFTYLEINPLVTKDGVYVLDLAAKVDATADYICKVKWGDIEFPPFPF  
 REAYPEEAYIADLDAKSGASLKLTLNPKGRIDWTMVAGGGASVVYSDTICDLGGVNELAN  
 YGEYS GAPSEQQTYDYAKTILSLMTREKHPDGKILIGGSIANFTNVAATFKGIVRAIRDYQ  
 GPLKEHEVTIFVRRGGPNYQEGLRVMGEVGKTTGIPIHVFGTETHMTAIVGMALGHRPIP  
 55 QPPTAAHTANFLLNASGSTSTPAPSRTASFSESRADEVAPAKKAKPAMPQDSVPSRSLQG  
 KSTTLFSRHTKAIVWGMQTRAVQGMDFDYVCSRDEPSVAAMVYPFTGDHKQKFYWGHI  
 KEILIPVFKNMADAMRKHPEVDVLINFASLRSAYDSTMETMNYAQIRTIHAEIGPEALTRK  
 LIKKADQKGVTHGPATVGGIKPGCFKJGNTGGMLDNILASKLYRPGSVAYVSRSGGMSNEL  
 NNIISRTTDGVYEGVAIGGDRYPGSTFMDHVLRYQDTPGVKMIVVLGEIGGTEEYKICRGIK  
 60 EGRLTKPIVCWCIGTCATMFSSEVQFGHAGACANQASETAVAKNQALKEAGVFVPRSFDE  
 LGEIQSVYEDLVANGVIVPAQEVPPPTVPMDYWARELGLIRKPASFMTSICDERGQELIY  
 AGMPITEVFKEEMGIGGVLLWFQKRLPKYSCQFIEMCLMVTADHGPAVSGAHNTIICAR  
 AGKDLVSSLTSGLLTIGDRFGGALDAAAKMFSAFDSGIIPMEFVNKMKKEGKLIMGIGHR

VKSINNPDMRVQILKDYVRQHFPATPLLDYALEVEKITTSSKKPNLILNVDGLIGVAFVDMIL  
RNCGSFTREEADEYIDIGALNGIFVLGRSMGFIGHYLDQKRLKQGLYRHPWDDISYVLPEH  
MSM

5

Table 4. Clustal W, Protein Domains, Cellular Location and Locus

The following is an alignment of the protein sequences of the human (CG142427-01), rat (J05210) and mouse (AF332052) versions of the ATP Citrate Lyase.

10

Multiple Alignment:

J05210	1	MDAKAISENTQKELLVHYICTTCAIQHFFYARVTPCTOWHLLQCHPWLLSQSLVYKFF	60
CG14247-01	1	MDAKAISENTQKELLVHYICTTCAIQHFFYARVTPCTOWHLLQCHPWLLSQSLVYKFF	60
AF332052	1	MDAKAISENTQKELLVHYICTTCAIQHFFYARVTPCTOWHLLQCHPWLLSQSLVYKFF	60
J05210	61	QLIHPPQGLGLVQVNLSDGVKSWLFPLOHEATVGLADGFLHFLIEPFFVHQQSEFF	120
CG14247-01	61	QLIHPPQGLGLVQVNLSDGVKSWLFPLOHEATVGLADGFLHFLIEPFFVHQQSEFF	120
AF332052	61	QLIHPPQGLGLVQVNLSDGVKSWLFPLOHEATVGLADGFLHFLIEPFFVHQQSEFF	120
J05210	121	SCVATRESCVPLFHHSDGVVDSSTLTGAKLISVSGEHLNAECIKFHLIHHAPPEKFI	180
CG14247-01	121	SCVATRESCVPLFHHSDGVVDSSTLTGAKLISVSGEHLNAECIKFHLIHHAPPEKFI	180
AF332052	121	SCVATRESCVPLFHHSDGVVDSSTLTGAKLISVSGEHLNAECIKFHLIHHAPPEKFI	180
J05210	181	LAKFICQLPHFYELLYFTYLEINPLVYTHDGVYILDLAAHPDAAATVYTHRWQDIEFF	240
CG14247-01	181	LAKFICQLPHFYELLYFTYLEINPLVYTHDGVYILDLAAHPDAAATVYTHRWQDIEFF	240
AF332052	181	LAKFICQLPHFYELLYFTYLEINPLVYTHDGVYILDLAAHPDAAATVYTHRWQDIEFF	240
J05210	241	PFQREAYFEENVIADLDANQASCLELTLINPQPIWTHFAAGGACETTCITNTISQDE	300
CG14247-01	241	PFQREAYFEENVIADLDANQASCLELTLINPQPIWTHFAAGGACETTCITNTISQDE	300
AF332052	241	PFQREAYFEENVIADLDANQASCLELTLINPQPIWTHFAAGGACETTCITNTISQDE	300
J05210	301	LANNVSGQAPSEGTVOYARTILSLMTRKHHQKILLIISQSLASFTTAASTFGKIFA	360
CG14247-01	301	LANNVSGQAPSEGTVOYARTILSLMTRKHHQKILLIISQSLASFTTAASTFGKIFA	360
AF332052	301	LANNVSGQAPSEGTVOYARTILSLMTRKHHQKILLIISQSLASFTTAASTFGKIFA	360
J05210	361	IRVQVQSLIEHEHTIFVRRQDPNYQGLAVHGEVQNTTQIIFHPFQTEIHTMTAIVQIAWA	420
CG14247-01	361	IRVQVQSLIEHEHTIFVRRQDPNYQGLAVHGEVQNTTQIIFHPFQTEIHTMTAIVQIAWA	420
AF332052	361	IRVQVQSLIEHEHTIFVRRQDPNYQGLAVHGEVQNTTQIIFHPFQTEIHTMTAIVQIAWA	420
J05210	421	PAPHPQPTAANTAFILNASQSTTPAPSPATSPSPSPADAPAFKKAAPAMHPDQSP	479
CG14247-01	421	PAPHPQPTAANTAFILNASQSTTPAPSPATSPSPSPADAPAFKKAAPAMHPDQSP	479
AF332052	421	PAPHPQPTAANTAFILNASQSTTPAPSPATSPSPSPADAPAFKKAAPAMHPDQSP	479
J05210	480	SPSFLQKQKATLSEPHTKAIVWQHNTPAWQMLRQGVVQSPDEPQWAMYPPTJQHKKQ	539
CG14247-01	480	SPSFLQKQKATLSEPHTKAIVWQHNTPAWQMLRQGVVQSPDEPQWAMYPPTJQHKKQ	539
AF332052	480	SPSFLQKQKATLSEPHTKAIVWQHNTPAWQMLRQGVVQSPDEPQWAMYPPTJQHKKQ	539
J05210	540	FYVQHICELIPFFNNHACAMKHPPEVDVLINFALESAVGTTHETMNYAQRTIAIIAEQ	599
CG14247-01	540	FYVQHICELIPFFNNHACAMKHPPEVDVLINFALESAVGTTHETMNYAQRTIAIIAEQ	599
AF332052	540	FYVQHICELIPFFNNHACAMKHPPEVDVLINFALESAVGTTHETMNYAQRTIAIIAEQ	599
J05210	600	IFPAITRELIEACQEQVTIIGPATVQGGHPGCFHIOHTQMLCHILASGLYPPQSPAY	659
CG14247-01	600	IFPAITRELIEACQEQVTIIGPATVQGGHPGCFHIOHTQMLCHILASGLYPPQSPAY	659
AF332052	600	IFPAITRELIEACQEQVTIIGPATVQGGHPGCFHIOHTQMLCHILASGLYPPQSPAY	659
J05210	660	SRSDQSHNELNHIISPTTQGVYEDVAIQGDRYFQSTFMCHPLVQQTTPQHEMIYVLSEID	719
CG14247-01	660	SRSDQSHNELNHIISPTTQGVYEDVAIQGDRYFQSTFMCHPLVQQTTPQHEMIYVLSEID	719
AF332052	660	SRSDQSHNELNHIISPTTQGVYEDVAIQGDRYFQSTFMCHPLVQQTTPQHEMIYVLSEID	719
J05210	720	STEEVYICQVILEQRLTKPQVQWCISQATMFCSEVGFQHGAGACACETAVAKHQALL	779
CG14247-01	720	STEEVYICQVILEQRLTKPQVQWCISQATMFCSEVGFQHGAGACACETAVAKHQALL	779
AF332052	720	STEEVYICQVILEQRLTKPQVQWCISQATMFCSEVGFQHGAGACACETAVAKHQALL	779
J05210	780	EAQVFFPFFQELAEIISQGVYEDVADQAVPAGQVPPPTPHMYSWAPRLGLIPNAPSE	839
CG14247-01	780	EAQVFFPFFQELAEIISQGVYEDVADQAVPAGQVPPPTPHMYSWAPRLGLIPNAPSE	839
AF332052	780	EAQVFFPFFQELAEIISQGVYEDVADQAVPAGQVPPPTPHMYSWAPRLGLIPNAPSE	839
J05210	840	STSLQDQPGGELLVAGHPITEVFEEHNDIGVGLLWFQKPLPEYSCQPIEMCLMTAQH	899
CG14247-01	840	STSLQDQPGGELLVAGHPITEVFEEHNDIGVGLLWFQKPLPEYSCQPIEMCLMTAQH	899
AF332052	840	STSLQDQPGGELLVAGHPITEVFEEHNDIGVGLLWFQKPLPEYSCQPIEMCLMTAQH	899
J05210	900	SPAVQSHHITICAPAKHGLVQGLTQGLTIQEPFGAGLGAAMHFKAPDQGLIPIHEFF	959
CG14247-01	900	SPAVQSHHITICAPAKHGLVQGLTQGLTIQEPFGAGLGAAMHFKAPDQGLIPIHEFF	959
AF332052	900	SPAVQSHHITICAPAKHGLVQGLTQGLTIQEPFGAGLGAAMHFKAPDQGLIPIHEFF	959
J05210	960	MMHHEGMLIMQHPVCSLNNPDMRVQILKDYVRQHFPATPLLDYALEVEKITTSSKKFN	1019
CG14247-01	960	MMHHEGMLIMQHPVCSLNNPDMRVQILKDYVRQHFPATPLLDYALEVEKITTSSKKFN	1019
AF332052	960	MMHHEGMLIMQHPVCSLNNPDMRVQILKDYVRQHFPATPLLDYALEVEKITTSSKKFN	1019
J05210	1020	LILNVQDQVAFPMHLENQSFTRERAGEYVIGALNIEVYLSRQHPFQHYLLQKRL	1079
CG14247-01	1020	LILNVQDQVAFPMHLENQSFTRERAGEYVIGALNIEVYLSRQHPFQHYLLQKRL	1079
AF332052	1020	LILNVQDQVAFPMHLENQSFTRERAGEYVIGALNIEVYLSRQHPFQHYLLQKRL	1079
J05210	1080	QGLYRHPWDDISYVLPEHMSM	1100
CG14247-01	1080	QGLYRHPWDDISYVLPEHMSM	1100
AF332052	1080	QGLYRHPWDDISYVLPEHMSM	1100

**Human ATP Citrate Lyase**

1105 amino acids; 121 kd

Locus: **17q12-q21**

Intracellular (Cytoplasmic)

5

In addition to the human version of the ATP Citrate Lyase identified as being differentially expressed in the experimental study, other variants have been identified by direct sequencing of cDNAs derived from many different human tissues and from sequences in public databases. No splice-form variants have been identified at CuraGen  
 10 whereas several amino acid-changing cSNPs were identified. These are found below. The preferred variant of all those identified, to be used for screening purposes, is CG142427-01.

Table 5: The variants of the human ATP Citrate Lyase obtained from direct cloning and/or public databases

15

DNA Position	Strand	Alleles	AA Position	AA Change	public SNP #
363	Plus	A:G	<b>75</b>	Asn=>Asp	rs1058875
665	Plus	A:C	<b>175</b>	Glu=>Asp	rs2304497
2318	Plus	G:A	<b>726</b>	Lys=>Lys	rs1802731
2377	Plus	G:A	<b>746</b>	Gly=>Glu	rs1802730
2756	Plus	C:T	<b>873</b>	Leu=>Leu	rs2277697
3308	Plus	C:G	<b>1056</b>	Ala=>Ala	Rs1802732

**Biochemistry and Cell Line Expression**

20

The following summarizes the biochemistry surrounding the human ATP Citrate Lyase enzyme: ATP Citrate Lyase catalyzes the conversion of Citrate plus CoA in the presence of ATP into orthophosphate + Acetyl CoA + Oxaloacetate with a release of ADP. Acetyl CoA can then be used as a substrate for Fatty Acid synthesis.

Cell lines expressing the ATP Citrate Lyase enzyme can be obtained from the RTQ-  
 25 PCR results shown above. These and other ATP Citrate Lyase enzyme expressing cell lines could be used for screening purposes.

**Findings:**

An inhibitor to ATP Citrate Lyase will force Acetyl CoA to be produced by  
 30 alternative pathways, thus decreasing the available pool for fatty acid and triglyceride

synthesis. The decreased pool of Acetyl CoA will cause a down-regulation of the Cholesterol biosynthetic pathway preventing excess production of LXRa ligands

Taken in total, the data indicates that an inhibitor of the human ATP Citrate Lyase enzyme would be beneficial in the treatment of obesity and/or diabetes.

5

Sequences: The sequence of Acc. No. CG142427-01 is an *In silico* prediction based on sequences available in CuraGen's proprietary sequence databases or in the public human sequence databases, and provided either the full length DNA sequence, or some portion thereof.

10

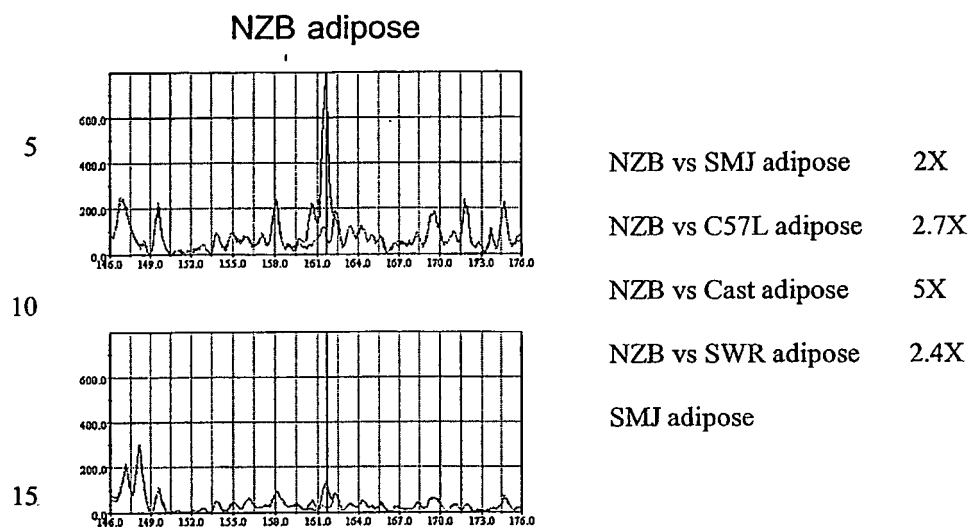
SPECIES #1. A gene fragment of the mouse ATP Citrate Lyase was initially found to be up-regulated by 2 fold in the adipose tissues of the NZB mouse relative to the SMJ mouse strain using CuraGen's GeneCalling™ method of differential gene expression. Similar results were found in adipose in NZB vs C57L, Cast and SWR mouse strains (All were up-regulated; 2.7x, 5x, and 2.4x respectively). A differentially expressed mouse gene fragment migrating, at approximately 161.7 nucleotides in length (Figures 1A and 1B. - vertical line) was definitively identified as a component of the mouse ATP Citrate Lyase cDNA (in the graphs, the abscissa is measured in lengths of nucleotides and the ordinate is measured as signal response). The method of competitive PCR was used for conformation of the gene assessment. The chromatographic peaks corresponding to the gene fragment of the rat ATP Citrate Lyase are ablated when a gene-specific primer (see below) competes with primers in the linker-adaptors during the PCR amplification. The peaks at 161.7 nt in length are ablated in the sample from both the NZB and SMJ mice.

25

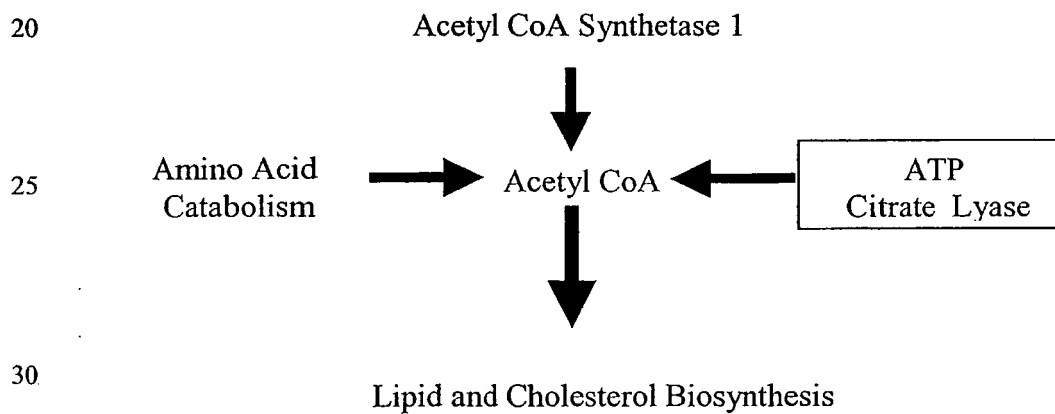
The direct sequence of the 65 nucleotide-long gene fragment and the gene-specific primers used for competitive PC are indicated on the complete cDNA sequence of the ATP Citrate Lyase and shown below in bold. The gene-specific primers at the 5' and 3' ends of the fragment are in bold.

30

35



**Figure 2. Schematic Showing the Role of ATP Citrate Lyase in Lipid and Cholesterol Biosynthesis.**



## F. NOV16a – Human Serine Dehydratase – CG142631-01

### Discovery Process

The following sections describe the study design(s) and the techniques used to identify the Serine Dehydratase - encoded protein and any variants, thereof, as being suitable as diagnostic markers, targets for an antibody therapeutic and targets for a small molecule drugs for obesity and/or diabetes.

#### Studies:

MB.01: Insulin Resistance in rat

#### Study Statements:

MB.01: The spontaneously hypertensive rat (SHR) is a strain exhibiting features of the human Metabolic Syndrome X. The phenotypic features include obesity, hyperglycemia, hypertension, dyslipidemia and dysfibrinolysis. Tissues were removed from adult male rats and a control strain (Wistar – Kyoto) to identify the gene expression differences that underlie the pathologic state in the SHR and in animals treated with various anti-hyperglycemic agents such as troglitazone. Tissues included sub-cutaneous adipose, visceral adipose and liver.

#### Species #1 Rat Strains SHR

#### Serine Dehydratase:

Serine dehydratase catalyzes the PLP-dependent alpha, beta-elimination of L-serine to pyruvate and ammonia. It is one of three enzymes that are regarded as metabolic exits of the serine-glycine pool. Serine dehydratase is found predominantly in the liver.

Table 1. Competitive PCR Primer for the Human Serine Dehydratase

Confirmatory Result – Human Serine Dehydratase (Discovery Study MB.01):

(Identified fragment from 221 to 545 in *italic*. band size: 325)

```

1      GCTTTATAAACATATATATATTAATTTTATTACAAAGTGCATATTATAAACATGGATAAAGGAGGGTGGGCC
81     ACTGTCAGGCGGACGCCACCCAGCCTACTCAGGGGTGCTGGTGACCCCTCAGGGTGGCCAGGGCAGCAGATATCAC
161    TTGAGTAGCTCATTCAGGCCAGCTGTGCCTTGAGTGCCCTGCAGCTGTGCCAGGCTGATGTTGCTGCCACCACACAAT
241    GACAACCAGCGAGGCCAGTGGGGTTTGAGTCGGGGCTCAGCCTGCAGCCTGCACACCACCCGCTGTACACTGCAGCCA
321    GGGCAGCGCCACACGCGGGCTCCACCAGGATCTTCTCATCGTCTACGAACTTCTCGATAGCAGTCACAGCCTCCTGGTCT
401    GAGATGACCTCAGAGAAAATGGGGTGTTCGTAAACAGCTTCAGGGTCTGTGCCCCACAGTGTTCACACCCAAGGCCTT

```

481 GGCAACACTGGTGATCTTGGGCAGGGTGACCAAGCTTCTCTTCTGACGGCAGGGTGAAGCTGTGGGCGCCGAAGGTCT  
 561 CCATGGCGATGATGGGCACATCTCCAGCCCACTCCCGCAGCCCTGGACCACTCCGCACAGCAGGCCCTCCACCGCCC  
 641 ACAGACAGCACAAATGGCCCCGGGCTTGGCGCTCAGTGTCTCTTTCAGCTCCTTCACAAGGGAAGTGTGGCCTTCCAGAT  
 721 GAGAGGGTTCATCGAAGGGGAGATGTACACCCAACTGGGTGTCTCTTCCAGAGCCTTGGCCAGTTGGATGGCCTCAT  
 5 CCAGCATCTCTCCCACTTCAACTGTGGCCCTTCGTTCTTCAGCCGCTCAATGGTGAGGGCAGGTGTGGTGTGGC  
 801 ACAACAATAGTGGCTGGGAGGCCAGCCTCCTGGCAGCATAGGCAGTCCCATGCCCGCTTGGCCGCTGAAGAGCAGAC  
 881 GAAATGTTTACAGCCTGTCTTGCCTTCATCTTGACAGATGCCCAATGCCTCGGATCTTGAAGGAGCCAGAGGGCTGAG  
 961 AGCTGTCCATCTTAAGGAACACACTAGTGCAGGCCACTTTGGACAATGCCATGCTGTACGTAGTGGGTCTTACGTGC  
 1041 AGGGACTCCTGGGCAGCCATGGCATGTAGCTTTGAAGGTGGATCCTCTGTCTCAGTCTCCCAATTGCTGGGATCACAG  
 1121 GTATGCCCGCCGCACCCGGCACAGGAGGAGCTGGACAGAGCGAGCGAGAAGGGTAGATTTTGTCTGTCTGGGAGAG  
 1201 TGGAAAGT  
 1281

15 Table 2. Nucleotide and protein sequence of Human Serine Dehydratase, CG142631-01

CCTTCTCTCGTGGGCTATCTACTCAGTTGATCCCTCCCTCGCTGGCTTGGCTCTGACTCCTG  
 CTCAGACCCATCACCTTTGCCGGGGAATGATGTCTGGAGAACCCCTGCACGTGAAGACCCCC  
 20 ATCCGTGACAGCATGGCCCTGTCCAAATGGCCGGCACCAGCGTCTACCTCAAGATGGACAG  
 TGCCCAGCCCTCCGGCTCCTTCAAGATCCGGGGCATTGGGCACCTCTGCAAGAGGTGGGCCA  
 AGCAAGGCTGTGCACATTTTGTCTGCTCCTCGGCGGGCAACGCAGGCATGGCGGCTGCATAT  
 GCGGCCAGGCAACTCGGCGTCCCCGCCACCATCGTAGTGCCCGGCACCACACCTGCTCTCA  
 CCATTGAGCGCCTCAAGAATGAAGGTGCCACATGCAAGGTGGTGGGTGAGTTATTGGATGAA  
 GCCTTCGAGCTGGCCAAGGCCCTAGCGAAGAACAACCCGGGTTGGGTCTACATTCCCCCTT  
 25 TGATGACCCCTCATCTGGGAAGGCCACGCTTCCATCGTGAAAGAGCTGAAGGAGACACTGT  
 GGGAAAAGCCGGGGGCCATCGCGCTGTCAAGTGGGCGGCGGGGGCCTGCTGTGTGGAGTGG  
 TCCAGGGGCTGCAGGAGTGTGGCTGGGGGGACGTGCCTGTCTCATCGCCATGGAGACTTTTGGT  
 GCCCACAGCTTCCACGCTGCCACCACCGCAGGCAAACTTGTCTCCCTGCCCAAGATCACCAG  
 GTTTGCCAAGGCCCTGGGCGTGAAGACTGTGGGGTCTCAGGCCCTGAAGCTGTTTCAGGAAC  
 30 ACCCCATTTTCTCTGAAGTTATCTCGGACCAGGAGGCTGTGGCCGCCATTGAGAAGTTCGTGG  
 ATGATGAGAAGATCCTGGTGGAGCCCGCCTGGGGCGCAGCCCTGGCCGCTGTCTATAGCCAC  
 GTGATCCAGAAGCTCCAACCTGGAGGGGAATCTCCGAACCCCGCTGCCATCCCTCGTGGTCAT  
 CGTCTGCGGGGGCAGCAACATCAGCCTGGCCAGCTGCGGGCGCTCAAGGAACAGCTGGGC  
 ATGACAAATAGGTTGCCCAAGTGAGGACGGACCCCTTACCGATCTGTGCTCTCCTAGCCCAAG  
 35 AGACCCCTGGAGGGGCTGGAGTTTATCCAGCGCCTCGTCGTATGTTTGGCTGAGCACCTGTG  
 GCCCTGGGTGCAGGTTAACTTCTTGTATCAGGAGCCCACTATGCAGAGGCCAAAGGTGGGC  
 AGCCAGCGAGGCTATGAATTGGACCTTTTGGTATCTGTGTGACTGCTCTGTGCCATCCTTA  
 GCCAATTGCTGGCGTGACAAGTGCCACAAGTAACACACCAGGTACCCAGAGCAGGGTGGAG  
 40 CAGGAGAGACCTGAATCACAGCAGTGAGG

Table 3. ORF Start: 90 ORF Stop: 1074 Frame: 3

#### Human Serine Dehydratase Protein Sequence:

**CG142631-01-prot** 328 aa

MMSGEPLHVKTPIRDSMALSKMAGTSVYLKMDSAQPSGSFKIRGIGHFCKRWAKQGCAHF  
 VCSSAGNAGMAAAYARQLGVPATIVPGTTPALTIERLKNEGATCKVVGELLDEAFELA  
 KALAKNNPGWVYIPFDDPLIWEGHASIVKELKETLWEKPGAIALSVGGGGLLCGVVQGL  
 QECGWGDVPVIAMETFGAHSFHAATTAGKLVSLPKITSVAKALGVKTVGSQALKLFQEHF  
 IFSEVISDQEAVAIEKFVDDEKILVEPAWGAAALAAVYSHVIQKLQLEGNLRTPLPSLVV  
 IVCGGSNISLAQLRALKEQLGMTNRLPK

Table 4. Clustal W, Protein Domains, Cellular Location and Locus

The following is an alignment of the protein sequences of the human (CG142631-01), rat and mouse versions of the Serine Dehydratase.

Multiple Alignment:	
10	<div> <div> DWR TT_Rat_SDH AAH11259_Mouse_SDH CG142631-01-SDH </div> <div> 1 1 1 </div> <div> MAAQESLHVKTPLRDSMALSI MAAQESLHVKTPLRDSMALSK MMSGEP LHVKTPLRDSMALSE </div> <div> 1 1 1 </div> <div> 60 60 60 </div> </div>
15	<div> <div> DWR TT_Rat_SDH AAH11259_Mouse_SDH CG142631-01-SDH </div> <div> 61 61 61 </div> <div> VSSAONAGMATAYAAARRLG VSSAONAGMATAYAAARRLG VSSAONAGMATAYAAARRLG </div> <div> 61 61 61 </div> <div> 120 120 120 </div> </div>
20	<div> <div> DWR TT_Rat_SDH AAH11259_Mouse_SDH CG142631-01-SDH </div> <div> 121 121 121 </div> <div> KALEKNNPQWVYISPFDDPLI KALEKNNPQWVYISPFDDPLI KALEKNNPQWVYISPFDDPLI </div> <div> 121 121 121 </div> <div> 180 180 180 </div> </div>
25	<div> <div> DWR TT_Rat_SDH AAH11259_Mouse_SDH CG142631-01-SDH </div> <div> 181 181 181 </div> <div> REVGVWELVPIIAMETFGAHS REVGVWELVPIIAMETFGAHS REVGVWELVPIIAMETFGAHS </div> <div> 181 181 181 </div> <div> 240 240 240 </div> </div>
30	<div> <div> DWR TT_Rat_SDH AAH11259_Mouse_SDH CG142631-01-SDH </div> <div> 241 241 241 </div> <div> IFSEVISDQEAVALIEKFVDD IFSEVISDQEAVALIEKFVDD IFSEVISDQEAVALIEKFVDD </div> <div> 241 241 241 </div> <div> 300 300 300 </div> </div>
35	<div> <div> DWR TT_Rat_SDH AAH11259_Mouse_SDH CG142631-01-SDH </div> <div> 301 301 301 </div> <div> VCGGSNTSLAQLRALHAEGL VCGGSNTSLAQLRALHAEGL VCGGSNTSLAQLRALHAEGL </div> <div> 301 301 301 </div> <div> 327 313 328 </div> </div>

Human Serine Dehydratase  
328 amino acids; 34 kd  
Locus: 12  
Intracellular

In addition to the human version of the Serine Dehydratase identified as being differentially expressed in the experimental study, other variants have been identified by direct sequencing of cDNAs derived from many different human tissues and from sequences in public databases. No splice-form variants have been identified at CuraGen whereas several amino acid-changing cSNPs were identified. These are found below. The preferred variant of all those identified, to be used for screening purposes, is CG142631-01.

Table 5. The variants of the human Serine Dehydratase obtained from direct cloning and/or public databases

DNA Position	Strand	Alleles	AA Position	AA Change	public SNP #
777	Plus	G:T	230	Ala => Ser	rs1050062

### Biochemistry:

- 5           The following illustrations summarizes the biochemistry surrounding the human Serine Dehydratase enzyme. L-Serine is converted to Pyruvate by pyridoxal phosphate requiring Serine Dehydratase with the release of ammonia as a by product. Pyruvate is a primary substrate in the process of gluconeogenesis. Cell lines expressing the Serine Dehydratase enzyme can be obtained from the RTQ-PCR results shown above. These and  
10 other Serine Dehydratase enzyme expressing cell lines could be used for screening purposes.

### Findings:

- 15 Serine Dehydratase (SDH) is critical for gluconeogenesis. In models of Diabetes SDH is up-regulated and in studies utilizing TZDs expression of SDH is down-regulated. An inhibitor of this enzyme would decrease glucose production. By improving daily blood glucose levels and maintaining HbA1c at or below 7.5 may prevent many diabetic complications.
- 20 Taken in total, the data indicates that an inhibitor of the human Serine Dehydratase enzyme would be beneficial in the treatment of obesity and/or diabetes.

### Sequences

- 25 The sequence of Acc. No. CG142631-01 is an *In silico* prediction based on sequences available in CuraGen's proprietary sequence databases or in the public human sequence databases, and provided either the full length DNA sequence, or some portion thereof.

5

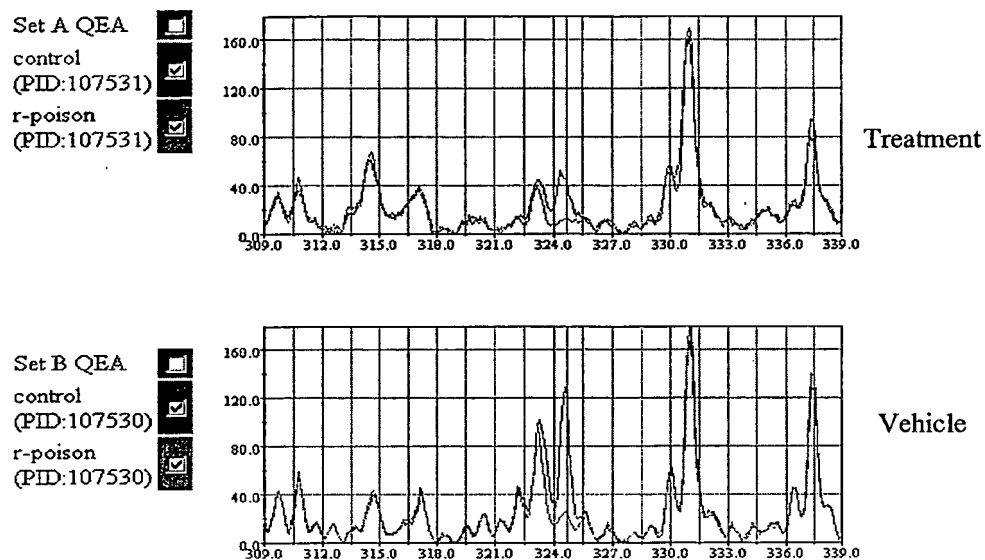
10

Treatment

Vehicle

Figures 1A and 1B. Differentially Expressed Gene Fragment from Rat Serine Dehydratase.

MB01: Troglitazone LD10 vs 0.02% DMSO WKY/72 hr -4



**G. NOV53a – Human Plasma Kallikrein – CG56155-01****Discovery Process**

The following sections describe the study design(s) and the techniques used to identify the Plasma Kallikrein - encoded protein and any variants, thereof, as being suitable as diagnostic markers, targets for an antibody therapeutic and targets for a small molecule drugs for Obesity and Diabetes.

MB.01: Metabolic Syndrome X in Rat

MB.04: Mouse Obesity

10

**Study Statements:**

MB.01 The spontaneously hypertensive rat (SHR) is a strain exhibiting features of the human Metabolic Syndrome X. The phenotypic features include obesity, hyperglycemia, hypertension, dyslipidemia and dysfibrinolysis. Tissues were removed from adult male rats and a control strain (Wistar – Kyoto) to identify the gene expression differences that underlie the pathologic state in the SHR and in animals treated with various anti-hyperglycemic agents such as troglitazone. Tissues included sub-cutaneous adipose, visceral adipose and liver.

MB.04 A large number of mouse strains have been identified that differ in body mass and composition. The AKR and NZB strains are obese, the SWR, C57L and C57BL/6 strains are of average weight whereas the SM/J and Cast/Ei strains are lean. Understanding the gene expression differences in the major metabolic tissues from these strains will elucidate the pathophysiologic basis for obesity. These specific strains of rat were chosen for differential gene expression analysis because quantitative trait loci (QTL) for body weight and related traits had been reported in published genetic studies. Tissues included whole brain, skeletal muscle, visceral adipose, and liver.

30

Species #1 Rat Strains SHR, WKY

Species #2 Mouse Strains C57BL, Cast/Ei

**Plasma Kallikrein:**

Plasma Kallikrein (PK) has been shown to activate specifically plasminogen during adipose differentiation. Plasminogen activation, followed by fibrinolysis, has been

implicated in adipose differentiation by remodeling of the fibronectin-rich extracellular matrix of preadipocytes.

5

**Table 1. SPECIES #1 Rat Plasma Kallikrein Gene Fragment used for competitive PCR**  
(fragment from 1516 to 1658 in **bold**, band size: 143)

```

1035 TCCCCAAGAC TGCAAGGCAG AGGGGTGTAA ATGTTCCCTA AGGTATATCCA CGGATGGCTC
1095 TCCAACCTAGG ATCACCTATG AGGCACAGGG GAGCTCTGGT TATTCTCTGA GACTGTGTAA
1155 AGTTGTGGAG AGCTCTGACT GTACGACAAA AATAAATGCA CGTATTGTGG GAGGAACAAA
1215 CTCCTCTTTA GGAGAGTGGC CATGGCAGGT CAGCCTGCAA GTGAAGTTGG TTTCTCAGAA
1275 CCATATGTGT GGAGGGTCCA TCATTGGACG CCAATGGATA CTGACGGCTG CCCATTGCTT
1335 TGATGGGATT CCCTATCCAG ACGTGTGGCG TATATATGGC GGGATTCTTA ATCTGTGAGA
1395 GATTACAAAC AAAACGCCTT TCTCAAGTAT AAAGGAGCTT ATTATTCATC AGAAATACAA
1455 AATGTCAGAA GGCAGTTACG ATATTGCCTT AATAAGCTT CAGACACCGT TGAATTATAC
1515 TGAATTCCAA AAACCAATAT GCCTGCCTTC CAAAGCTGAC ACAAATACAA TTTATACCAA
1575 CTGCTGGGTG ACTGGATGGG GCTACACAAA GGAACGAGGT GAGACCCAAA ATATTCTACA
1635 AAAGGCAACT ATTCCTTGGT TACCAAATGA AGAATGCCAG AAAAAATATA GAGATTATGT
1695 TATAACCAAG CAGATGATCT GTGCTGGCTA CAAAGAAGGT GGAATAGATG CTTGTAAGGG
1755 AGATTCCGGT GGCCCTTAGT TTTGCAAAACA TAGTGAAGG TGGCAGTTGG TGGTATCAC
1815 CAGCTGGGGT GAAGGCTGTG CCGCAAGGA GCAACCAGGA GTCTACACCA AAGTTGCTGA
1875 GTACATTGAC TGGATATTGG AGAAGATACA GAGCAGCAAG GAAAGAGCTC TGGAGACATC
1935 TCCAGCATGA GGAGGCTGGG TACTGACGGG GAAGAGCCCA GCTGGCACCA GCTTTACCAC
1995 CTGCCCTCAA GTCCTACTAG AGCTCCAGAG TTCTCTTCTG CAAAATGTGG ATAGTGGTGT
2055 CTACCTCGCA TCCTTACCAT AGGATTAAAA GTCCAAATGT AGACACAGTT GCTAAAGACA
2115 GCGCCATGCT CAAGCGTGCT TCCT

```

(gene length is 2444, only region from 1035 to 2138 shown)

**Table 2. SPECIES #2. Mouse Plasma Kallikrein Gene Fragment used for competitive PCR**  
(fragment from 2807 to 2902 in **bold**, band size: 96)

```

2326 GTAAGGGAGA TTCCGGTGGC CCCTTAGTCT GTAAACACAG TGGACGGTGG CAGTTGGTGG
2386 GTATCACCAG CTGGGGTGAA GGCTGCGGCC GCAAGGACCA ACCAGGAGTC TACACCAAAG
2446 TTTCTGAGTA CATGGACTGG ATATTGGAGA AGACACAGAG CAGTGATGTA AGAGCTCTGG
2506 AGACATCTTC AGCCTGAGGA GGCTGGGTAC CAAGGAGGAA GAACCCAGCT GGCTTTACCA
2566 CCTGCCCTCA AGGCAAACTA GAGCTCCAGG ATTCTCGGCT GTAAAATGTT GATAATGGTG
2626 TCTACCTCAC ATCCGTATCA TTGGATTGAA AATTCAAGTG TAGATATAGT TGCTGAAGAC
2686 AGCGTTTGTG TCAAGTGTGT TTCTGCTCTT GAGTCACAGG AGCTCCAATG GGAGCATTAC
2746 AAGATCACC AAGCTTGTTA GAAAAGAGAA TGATCAAAGG GTTTTATTAG GTAATGAAAT
2806 GTCTAGATGT GATGCAATTG AAAAAAGAC CCCAGATTTT AGCACAGTCC TTGGGACCAT
2866 TTTCTAGTAA CTGTTGACTT TGGACCTCAG CAGATCTCAG AGTTACCTGT CCACCTCTGA
2926 CATTGTGTTA TTAGAGCCTG ATGCTATTCT TTCAAGTGA GCAAAAAAAA AAAAAA
2986 AAAAA

```

(gene length is 2990, only region from 2326 to 2990 shown)

45

**Table 3. Human Plasma Kallikrein Gene and Protein Sequence.**

```

>CG56155-01 2245 nt
AGAACAGCTTGAAGACGTTTCATTTTAAAGTGACAAGAGACTCACCTCCAAGAAGCAATT
GTGTTTTAGAAATGATTTTATTCAGCAAGCAACTTATTTTCATTTCTTGTGCTACAG
TTTCTGTGGATGTCGACTCAACTCTATGAAAACGCCCTTCTCAGAGGTGGGATGTAG
CTTCCATGTACACCCCAATGCCAATACTGCCAGATGAGGTGCACATTCACCCCAAGGT
GTTTGCTATTCACTTTCTTCCAGCAAGTTCAATCAATGACATGGAGAAAAGGTTTGGTT
GCTTCTTGAAAGATAGTGTACAGGAACCCCTGCCAAAAGTACATCGAACAGGTGCAGTTT
CTGGACATTCCTTGAAGCAATGTGGTTCATCAATAAGTGCTTGCATCGAGACATTTATA
AAGGAGTTGATATGAGAGGAGTCAATTTAATGTGTCTAAGGTTAGCAGTGTGAAGAAT

```

5. GCCAAAAAGGTGCACCAATAACATTGCTGCCAGTTTTTTCATATGCCACGCAACAT  
 TTCAAGGCAGAGTACCGGAACAATTGCCTATTAAAGTACAGTCCCGGAGGAACACCTA  
 CCGCTATAAAGGTGCTGAGTAACCTGGAATCTGGATTCTCACTGAAGCCCTGTGCCCTTT  
 CAGAAATTGGTTGCCACATGAACATCTTCCAGCATCTTGCCTTCTCAGATGTGGATGTTG  
 CCAGGGTCTCACTCCAGATGCTTTGTGTGTCGGACCATCTGCACCTATCACCCCAACT  
 GCCTCTTCTTTACATTCTATACAAATGTATGGAAAAATCGAGTCACAAAGAAATGTTGTGTC  
 TTCTTAAACATCTGAAAGTGGCACACCAAGTTCCTCTACTCCTCAAGAAAAACCATAT  
 CTGGATATAGCCTTTTAACTGCAAAAGAACTTTACCTGAACCCCTGCCATTCTAAAAATT  
 10. ACCCGGGAGTTGACTTTGGAGGAGAAATGAATGTGACTTTGTTAAAGGAGTGAATG  
 TTTGCCAAGAGACTTGCACAAAGATGATTCGCTGTCAGTTTTTTCACTTATTCTTTACTCC  
 CAGAAGACTGTAAGGAAGAGAAAGTGAAGTGTCTTAAGATTATCTATGGATGGTTCTC  
 CAACTAGGATTGCGTATGGGACACAAGGGAGCTCTGGTTACTCTTTGAGATTGTGTAACA  
 CTGGGGACAACCTGCTCTGCACAAAAACAAGCACACGCTTGTGGAGGAACAACT  
 15. CTTCTTGGGGAGAGTGGCCCTGGCAGGTGAGCCCTGCAGGTGAAGCTGACAGCTCAGAGGC  
 ACCGTGTGGAGGGTCACTCATAGGACACCAAGTGGGTCTCTACTGTGCCCACTGCTTTG  
 ATGGGCTTCCCCTGCAGGATGTTGGCGCATCTATAGTGGCATTTTAAATCTGTACAGACA  
 TTACAAAAGATACACCTTTCTCAGAAATAAAGAGATTATTATTCACCAAACTATAAAG  
 TCTCAGAAGGAATCATGATATCGCCTTGATAAACTCCAGGCTCCTTTGAATTACACTG  
 20. AATTCCAAAAACCAATATGCCTACCTTCCAAAGGTGACACAAGCACAAATTTATACCAACT  
 GTTGGGTAAACCGATGGGGCTTCTCGAAGGAGAAAGGTGAAATCCAAATATTCTACAAA  
 AGGTAAATATTCCTTTGGTAACAAATGAAGAATGCCAGAAAGATATCAAGATTATAAAA  
 TAACCCAAACCGATGGTCTGTGCTGGCTATAAAGAAGGGGAAAGATGCTTGTAAAGGAG  
 ATTCAAGTGGTCCCTTAGTTTGCAACACACCGAATGTGGCGTTTGGTGGGCATCACAA  
 GCTGGGGTGAAGGCTGTGCCCGCAGGGAGCAACCTGGTGTCTACACCAAGTCTGCTGAGT  
 25. ACATGGACTGGATTTTAGAGAAAAACACAGAGCAGTGATGGAAAAAGCTCAGATGCAGTCAC  
 CAGCATGAGAAGCAGTCCAGAGTCTAGGCAATTTTTACAACTGAGTCAAGTCAAAATTC  
 TGAGCCTGGGGGGTCCCTCATCTGCAAGCATGGAGAGTGGCATCTTCTTGCATCCTAAG  
 GACGAAAGACACAGTGCACCTCAGAGCTGCTGAGGACAATGTCTGCTGAAGCCCGCTTTCA  
 30. GCACGCCGTAACCGGGCTGACAAATGCGAGGTGCAACTGAGATCTCCATGACTGTGTG  
 TTGTGAAATAAAATGGTGAAAGATC

Table 4. Amino acid sequence for Human Plasma Kallikrein

35 ORF Start: 72 ORF Stop: 1986 Frame: 3

## Human Plasma Kallikrein Protein Sequence:

>CG56155-01-prot 638 aa  
 MILFKQATYFISLFATVSCGLTQLYENAFFRGGDVASMYTPNAQYQCMRCTFHPRLCLLF  
 SFLPASSINDMEKRFPCFLKDSVTGTLPKVHRTGAVSGHSLKQCGHQISACHRDIYKGV  
 MRGVNPNVSKVSSVEECQKRCINNIRCOFFSYATQTFHKAERYNNCLLKYSPPGTPTAIK  
 VLSNVEGFSGLKPCALSEIGCHMNIHQHAFSDVDVARVLTDPDAFVCRITICTYHPNCLFF  
 TFYTNVWKIESQRNVCLLKTSBSGTPSSSTPQENTISGYSLLTCKRTLPEPCHSKIYPGV  
 DFGGEELNVTFVKGVNVCQETCTKMI RCQFFYSLLPEDCKEEKCKFLRLSMDGSPTRI  
 AYGTCGSSGYSRLRLCNTGDNVCTTKTSTRI VGGTNSWGEWPQVSLQVKLTAQRHLCG  
 GSLIGHQVWLTAACHFDGLPLQDVWRIYSGILNLSDIKDTPFPSQIKEIIHQNYKVSEG  
 NHDIALIKLQAPLNYTEFQKPICLPSKGDSTIYTNWVTGNGFSKEKGEIQNILQKVNI  
 PLVTNEECQKRYQDYKITQRMVCAGYKEGGKDAKGDGSGPLVCKHNGMWRLVGTISWGE  
 GCARRBQPGVYTKVAEYMDWILEKTQSSDGKAMQSPA

40 Table 5. Clustal W, Protein Domains, Cellular Location and Locus

The following is an alignment of the protein sequences of the human (CG56155-01), rat and mouse versions of the Plasma Kallikrein.

PK\_rat M L F K O V G Y F V S L F A T V S C G C L S Q L Y A N T F F R G G D L A A I Y T P D A O H J O K I M D T F H P R C L L F  
 PK\_mouse M L F N R F G Y F V S L F A T V S C G C M T Q L Y K N T F F R G G D L A A I Y T P D A O Y C Q I M C T F H P R C L L F  
 CG56155-01 M L F K O A T Y F I S L F A T V S C G C L T Q L Y E N A F F R G G D L A S M I C T P N A O Y C O M R J T F H P R C L L F

PK\_rat S F L A V S P T K E T D K F P C F M K E S I T G T L P R I H R T G A I S G H S L K Q C G H O I S A C H Q D I Y E G L D  
 PK\_mouse S F L A V T P P K E T N K F P C F M K E S I T G T L P R I H R T G A I S G H S L K Q C G H O I S A C H R D I Y K G L D  
 CG56155-01 S F L P A S S I N D M E F F G C F L K E S M I G T L P K V H R T G A S G H S L K Q C G H O I S A C H R D I Y E G L D

PK\_rat M R G S N F N I S K T D S I E C Q E L C T N N I E C Q F F T Y A T K A F H R P E Y R K S P L L E R S S S O T P T S I K  
 PK\_mouse M R G S N F N I S K T D N I E C Q E L C T N N F E C Q F F T Y A T S A F Y R P E Y R K C L L H S A S O T P T S I K  
 CG56155-01 M R G Y N F N Y S E F S S V E C Q E R C T N N I R C Q F F S Y A T Q T F H K A E Y K N N E L L E Y S P G S T P T A I K

PK\_rat P F D N L V S G F S L K S C A L S E I G C P M D I F C H A F A D L N V S Q V T P D A Y V C R T I C T F H P H C L L F  
 PK\_mouse S A D N L V S G F S L K S C A L S E I G C P M D I F C H S A F A D L N V S Q V T P D A Y V C R T I C T F H P H C L L F  
 CG56155-01 Y E S H V E S G F S L K P C A L S E I G C H M N I F C H L A S D U D A R V T P D A Y V C R T I C T F H P H C L L F

PK\_rat F F Y T N E W E T E S Q R N V C F L E T S K S G E F S P P I I D E N A S G Y S L F P O K K A P P E P C H E K I Y S S V  
 PK\_mouse T F Y T N E W E T E S Q R N V C F L E T S K S G E F S P P I P Q E N A S G Y S L L T O R I E T E P E P C H S K I Y S S V  
 CG56155-01 T F Y T N Y W K I E S Q R N V C L L E T S E S G T E S S T P O E N F S G Y S L L T O R I E P C H S K I Y P S V

PK\_rat A F E G E E L N A F V Q G A L A C Q E T C T K T I R C Q F F T Y S L L P O D C K A E G C K C L R L S T D G S P T R I  
 PK\_mouse D F E G E E L N V T F V Q G A L V C Q E T C T K T I R C Q F F T Y S L L P O D C K E E G C K C L R L S T D G S P T R I  
 CG56155-01 D F G G E E L N V T F V K L V N V C Q E T C T K M I R C Q F F T Y S L L P E D C K E E K P K C F L R L S M A S P T R I

PK\_rat T Y E A D G S S G Y S L R L C K V E S S D C T T R I N A R I 7 G G T N S S L G E W P W O V S L Q 7 K L 7 S O N H M D U  
 PK\_mouse T Y G M D G S S G Y S L R L C K V D S P D C T T E I N A R I 7 G G T H A S L G E W P W O V S L Q 7 K L 7 S O T H L D G  
 CG56155-01 A M G T D G S S G Y S L R L C N T G D N S V C T T K T S T R I 7 G G T N S S W E W P W O V S L Q 7 A L T A J R H L D G

PK\_rat G S I I G R O W L T A A H C F L G P P Y P L V W H I Y G G I L N L S E I T N K T P E S S K E L I I H O K T K M S E G  
 PK\_mouse G S I I G R O W L T A A H C F L G P P Y P D V W E I Y G G I L S L S E T K E T P S S R K E L I I H O E Y K V S E G  
 CG56155-01 G S I I G H O W L T A A H C F L G P F L Q L V W H I Y S G I L N L S D T K E T P E S S Q K E L I I H O N K R V S E G

PK\_rat S Y D I A L I K L O T P L N Y T E F O K P I C L F S K A D T N T Y T N C W Y T G W G Y T K E Q E T O M I L O K A T I  
 PK\_mouse H Y D I A L I K L O T P L N Y T E F O K P I C L F S K A D T N T Y T N C W Y T G W G Y T K E Q E T O M I L O K A T I  
 CG56155-01 N H D I A L I K L O A P L N Y T E F O K P I C L F S K E L F S T Y T N C W Y T G W G S E E K Q E T J H I L O K V N I

PK\_rat P L V P N E E C Q K E Y R D Y V I T E Q M I C A G Y K E G S I D A C K G D S G G P L V C K H S G R W O L V G I T S W S E  
 PK\_mouse P L V P N E E C Q K E Y R D Y V I N E Q M I C A G Y K E G S I D A C K G D S G G P L V C K H S G R W O L V G I T S W S E  
 CG56155-01 P L V T N E E C Q K E Y R D Y K I T Q R M V C A G Y K E G S I D A C K G D S G G P L V C K H N J M W R L V G I T S W S E

PK\_rat G C A R K E O P G V Y T K Y A E Y I D W I L E K I T S S K E R A L E T S P A  
 PK\_mouse G C S R K E O P G V Y T K Y S E Y M D W I L E K T Q S S D V R A L E T S S A  
 CG56155-01 G C A R R E O P G V Y T K Y A E Y M D W I L E K T Q S S L G K A Q M Q S P A

- 5 Human Plasma Kallikrein  
 Locus: 4q35  
 Extracellular

10 In addition to the human version of the Plasma Kallikrein identified as being differentially expressed in the experimental study, other variants have been identified by direct sequencing of cDNAs derived from many different human tissues and from sequences in public databases. No splice-form variants have been identified at CuraGen whereas several amino acid-changing cSNPs were identified. These are found below. The preferred variant of all those identified, to be used for screening purposes, is CG56155-01.

**Table 6.** The variants of the human Plasma Kallikrein obtained from direct cloning and/or public databases

DNA Position	Strand	Alleles	AA Position	AA Change	public SNP #
499	Minus	A:G	143	Asn => Ser	
726	Minus	G:T	219	Val => Phe	
726	Minus	T:G	219	Val => Phe	
1212	Minus	T:G	381	Ser => Ala	
1272	Minus	T:G	401	Glu =>	
1832	Minus	C:T	587	Asn => Asn	
2073	Minus	G:A	0		
2073	Minus	A:G	0		

5

### Expression Profiles:

**Table 7.** CG56155-01: Plasma kallikrein - isoform1, submitted to study  
DDAT on 01/09/01 by sspaderna; clone status=FIS; novelty=Public;  
ORF start=72, ORF stop=1986, frame=3; 2245 bp.

10

Expression of gene CG56155-01 was assessed using the primer-probe set Ag1688, described in Table 7. Results of the RTQ-PCR runs are shown in Tables 8 and 9.

**Table 7.** Probe Name Ag1688

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-tcagaagggaatcatgatatcg-3'	22	1503	627
Probe	TET-5'-ccttgataaaactccaggctcctttga-3'-TAMRA	27	1525	628
Reverse	5'-tttggaaggtaggcatattgg-3'	21	1572	629

15 **Table 8.** Panel 1.3D

Tissue Name	Rel. Exp.(%) Ag1688, Run 147249266	Tissue Name	Rel. Exp.(%) Ag1688, Run 147249266
Liver adenocarcinoma	0.0	Kidney (fetal)	9.2
Pancreas	6.7	Renal ca. 786-0	0.0
Pancreatic ca. CAPAN 2	0.2	Renal ca. A498	1.7
Adrenal gland	1.8	Renal ca. RXF 393	0.0
Thyroid	3.8	Renal ca. ACHN	0.0
Salivary gland	1.5	Renal ca. UO-31	0.0
Pituitary gland	6.1	Renal ca. TK-10	0.0

Brain (fetal)	0.5	Liver	100.0
Brain (whole)	3.6	Liver (fetal)	99.3
Brain (amygdala)	3.3	Liver ca. (hepatoblast) HepG2	0.0
Brain (cerebellum)	0.4	Lung	1.3
Brain (hippocampus)	6.2	Lung (fetal)	1.8
Brain (substantia nigra)	1.0	Lung ca. (small cell) LX-1	0.0
Brain (thalamus)	2.1	Lung ca. (small cell) NCI-H69	0.0
Cerebral Cortex	6.3	Lung ca. (s.cell var.) SHP-77	0.8
Spinal cord	3.1	Lung ca. (large cell)NCI-H460	0.0
glio/astro U87-MG	0.0	Lung ca. (non-sm. cell) A549	0.2
glio/astro U-118-MG	0.0	Lung ca. (non-s.cell) NCI-H23	0.0
astrocytoma SW1783	0.0	Lung ca. (non-s.cell) HOP-62	0.0
neuro*; met SK-N-AS	0.2	Lung ca. (non-s.cl) NCI-H522	0.0
astrocytoma SF-539	0.0	Lung ca. (squam.) SW 900	0.2
astrocytoma SNB-75	0.1	Lung ca. (squam.) NCI-H596	0.0
glioma SNB-19	0.2	Mammary gland	2.9
glioma U251	1.2	Breast ca.* (pl.ef) MCF-7	0.0
glioma SF-295	0.0	Breast ca.* (pl.ef) MDA-MB-231	0.0
Heart (Fetal)	0.2	Breast ca.* (pl. ef) T47D	0.0
Heart	1.6	Breast ca. BT-549	0.0
Skeletal muscle (Fetal)	0.7	Breast ca. MDA-N	0.0
Skeletal muscle	1.2	Ovary	0.0
Bone marrow	0.5	Ovarian ca. OVCAR-3	0.2
Thymus	3.2	Ovarian ca. OVCAR-4	0.0
Spleen	1.0	Ovarian ca. OVCAR-5	0.3
Lymph node	2.9	Ovarian ca. OVCAR-8	0.0

Colorectal	0.8	Ovarian ca. IGROV-1	0.0
Stomach	3.3	Ovarian ca. (ascites) SK-OV-3	1.0
Small intestine	6.2	Uterus	1.4
Colon ca. SW480	0.0	Placenta	0.4
Colon ca.* SW620 (SW480 met)	0.0	Prostate	1.0
Colon ca. HT29	0.0	Prostate ca.* (bone met) PC-3	0.0
Colon ca. HCT-116	0.0	Testis	6.1
Colon ca. CaCo-2	0.2	Melanoma Hs688(A).T	0.4
CC Well to Mod Diff (ODO3866)	0.0	Melanoma* (met) Hs688(B).T	0.9
Colon ca. HCC-2998	0.2	Melanoma UACC-62	0.0
Gastric ca. (liver met) NCI-N87	4.4	Melanoma M14	0.0
Bladder	3.1	Melanoma LOX IMVI	0.0
Trachea	3.0	Melanoma* (met) SK-MEL-5	0.0
Kidney	6.8	Adipose	0.5

Table 9. Panel 5 Islet

Tissue Name	Rel. Exp.(%) Ag1688, Run 226587524	Tissue Name	Rel. Exp.(%) Ag1688, Run 226587524
97457_Patient-02go adipose	41.2	94709_Donor 2 AM - A_adipose	0.0
97476_Patient-07sk_skeletal muscle	9.9	94710_Donor 2 AM - B_adipose	0.0
97477_Patient-07ut_uterus	8.1	94711_Donor 2 AM - C_adipose	0.0
97478_Patient-07pl_placenta	0.0	94712_Donor 2 AD - A_adipose	11.4
99167_Bayer Patient 1	84.7	94713_Donor 2 AD - B_adipose	0.0
97482_Patient-08ut_uterus	2.4	94714_Donor 2 AD - C_adipose	29.1
97483_Patient-08pl_placenta	0.0	94742_Donor 3 U - A Mesenchymal Stem Cells	19.2
97486_Patient-	8.0	94743_Donor 3 U -	0.0

09sk_skeletal muscle		B_Mesenchymal Stem Cells	
97487_Patient-09ut_uterus	9.6	94730_Donor 3 AM - A_adipose	15.0
97488_Patient-09pl_placenta	0.0	94731_Donor 3 AM - B_adipose	37.9
97492_Patient-10ut_uterus	0.0	94732_Donor 3 AM - C_adipose	0.0
97493_Patient-10pl_placenta	0.0	94733_Donor 3 AD - A_adipose	39.2
97495_Patient-11go_adipose	0.0	94734_Donor 3 AD - B_adipose	11.4
97496_Patient-11sk_skeletal muscle	52.9	94735_Donor 3 AD - C_adipose	34.4
97497_Patient-11ut_uterus	35.8	77138_Liver_HepG2untreated	8.4
97498_Patient-11pl_placenta	10.5	73556_Heart_Cardiac stromal cells (primary)	0.0
97500_Patient-12go_adipose	0.0	81735_Small Intestine	100.0
97501_Patient-12sk_skeletal muscle	35.4	72409_Kidney Proximal Convoluted Tubule	9.9
97502_Patient-12ut_uterus	20.7	82685_Small intestine Duodenum	70.2
97503_Patient-12pl_placenta	0.0	90650_Adrenal_Adrenocortical adenoma	25.5
94721_Donor 2 U - A_Mesenchymal Stem Cells	0.0	72410_Kidney_HRCE	10.4
94722_Donor 2 U - B_Mesenchymal Stem Cells	0.0	72411_Kidney_HRE	7.2
94723_Donor 2 U - C_Mesenchymal Stem Cells	0.0	73139_Uterus_Uterine smooth muscle cells	0.0

### Biochemistry and Cell Line Expression

- 5 Plasma Kallikrein is a protease which is implicated in the conversion of plasminogen to the plasmin. Plasma Kallikrein activity was measured usually by spectrophotometric assays using artificial fluorescent peptide substrates. Plasma Kallikrein is commercially available enzyme with known inhibitors. The procedure of purification of Plasma Kallikrein from serum by affinity chromatography was described in literature. Cell lines expressing the

Plasma Kallikrein can be obtained from the RTQ-PCR results shown above. These and other Plasma Kallikrein expressing cell lines could be used for screening purposes.

**Rationale for use as a diagnostic and/or target for small molecule drugs and antibody  
5 therapeutics.**

1. Plasminogen activation, followed by fibrinolysis, is implicated recently in adipose differentiation by remodeling of the fibronectin-rich ECM of the preadipocytes. Knock out of the plasminogen gene in mouse lead to the reduction of fat deposit.
- 10 2. Plasma Kallikrein activates plasminogen, thus promoting adipose differentiation.
3. Plasma Kallikrein is significantly down-regulated in the liver of mice with the lean phenotype, which may cause disruption of the adipose differentiation ion this strain.
4. Taken in total, the data indicates that an inhibitor/antagonist of the human Plasma Kallikrein would be beneficial in the treatment of obesity.

15

SPECIES #1 A gene fragment of the rat Plasma Kallikrein was initially found to be down-regulated by 2 fold in MB.01 study in the liver of SHR rat relative to normal control rat strain using CuraGen's GeneCalling™ method of differential gene expression. Additionally, the expression of the enzyme was increased in the response to troglitazone treatment. A differentially expressed rat gene fragment migrating, at approximately 142.3 nucleotides in length (Figure 1a. - vertical line) was definitively identified as a component of the rat Plasma Kallikrein cDNA (in the graphs, the abscissa is measured in lengths of nucleotides and the ordinate is measured as signal response). The method of competitive PCR was used for conformation of the gene assessment. The electropherogram peaks  
20 corresponding to the gene fragment of the rat Plasma Kallikrein are ablated when a gene-specific primer (see below) competes with primers in the linker-adaptors during the PCR amplification. The peaks at 142.3 nt in length are ablated in the sample from both the SHR and control rats.

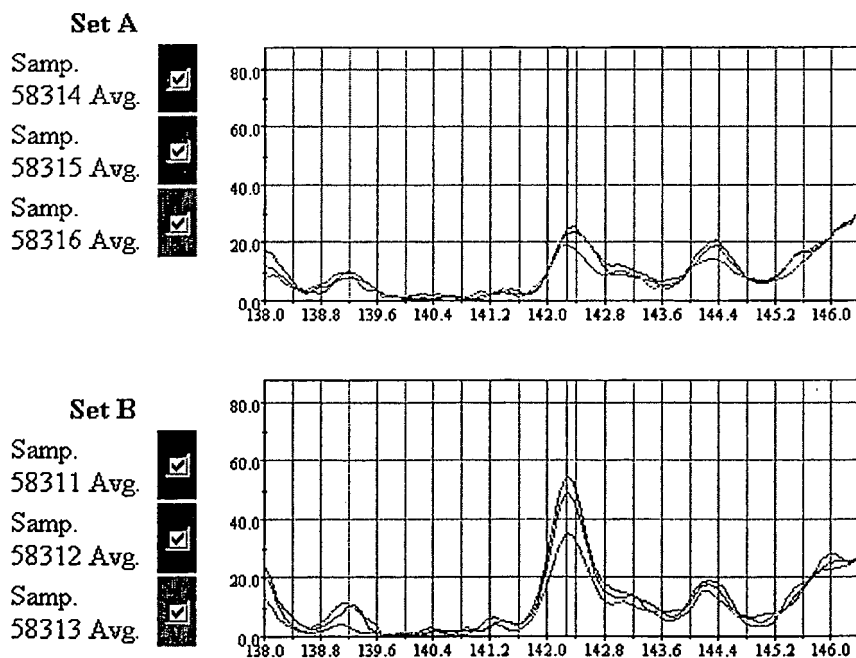
30 SPECIES #2 The gene fragments corresponding to the mouse Plasma Kallikrein were found to be down-regulated by 52.1 fold in liver tissues of normal mice relative to the lean mice. A differentially expressed mouse gene fragment migrating, at approximately 96 nucleotides in length (Figure 1a. - red vertical line) was definitively identified as a

component of the mouse Plasma Kallikrein cDNA by the method of competitive PCR. The electropherogramatic peaks corresponding to the gene fragment of the mouse Plasma Kallikrein are ablated when a gene-specific primer (see below) competes with primers in the linker-adaptors during the PCR amplification. The peaks at 96 nt in length are ablated in the sample from both the normal and lean mice.

The sequence of the nucleotide-long gene fragment and the gene-specific primers used for competitive PCR are indicated on the cDNA sequence of the Plasma Kallikrein and shown below in bold. The gene-specific primers at the 5' and 3' ends of the fragment are in color.

Figures 1A and 1B. Differentially Expressed Rat Plasma Kallikrein in Study MB.01.

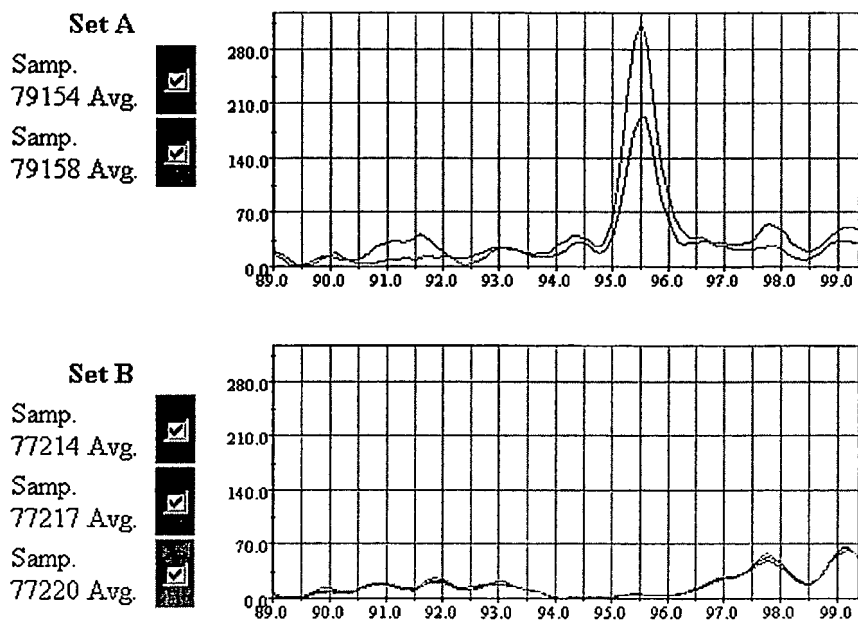
SPECIES #1



Figures 2A and 2B. Differentially Expressed Mouse Plasma Kallikrein in Study MB.04.

SPECIES #2

5.



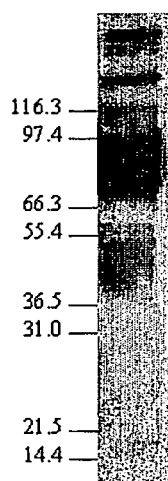
10

15

**Example F. CG56155-03 Expression data:**

**Construction of the mammalian expression vector pCEP4/Sec.** The oligonucleotide primers, pSec-V5-His Forward (CTCGTCCTCGAGGGTAAGCCTATCCCT AAC) and the pSec-V5-His Reverse (CTCGTCGGGCCCCTGATCAGCGGGTTTAAAC), were  
 5 designed to amplify a fragment from the pcDNA3.1-V5His (Invitrogen, Carlsbad, CA) expression vector. The PCR product was digested with XhoI and ApaI and ligated into the XhoI/ApaI digested pSecTag2 B vector (Invitrogen, Carlsbad CA). The correct structure of the resulting vector, pSecV5His, was verified by DNA sequence analysis. The vector pSecV5His was digested with PmeI and NheI, and the PmeI-NheI fragment was ligated  
 10 into the BamHI/Klenow and NheI treated vector pCEP4 (Invitrogen, Carlsbad, CA). The resulting vector was named as pCEP4/Sec.

**Expression of CG56155-03 in human embryonic kidney 293 cells.** A 0.4 kb BamHI-XhoI fragment containing the CG56155-03 sequence was subcloned into BamHI-XhoI  
 15 digested pCEP4/Sec to generate plasmid 1061. The resulting plasmid 1061 was transfected into 293 cells using the LipofectaminePlus reagent following the manufacturer's instructions (Gibco/BRL). The cell pellet and supernatant were harvested 72h post transfection and examined for CG56155-03 expression by Western blot (reducing conditions) using an anti-V5 antibody. Fig. 1 shows that CG56155-03 is expressed as a 74  
 20 kDa protein secreted by 293 cells.



**Fig. 1.** CG56155-03 protein secreted by 293 cells.

**PAGE INTENTIONALLY LEFT BLANK**

### **OTHER EMBODIMENTS**

- Although particular embodiments have been disclosed herein in detail, this has been done by way of example for purposes of illustration only, and is not intended to be limiting with respect to the scope of the appended claims, which follow. In particular, it is
5. contemplated by the inventors that various substitutions, alterations, and modifications may be made to the invention without departing from the spirit and scope of the invention as defined by the claims. The choice of nucleic acid starting material, clone of interest, or library type is believed to be a matter of routine for a person of ordinary skill in the art with knowledge of the embodiments described herein. Other aspects, advantages, and
  10. modifications considered to be within the scope of the following claims. The claims presented are representative of the inventions disclosed herein. Other, unclaimed inventions are also contemplated. Applicants reserve the right to pursue such inventions in later claims.

**CLAIMS**

What is claimed is:

1. An isolated polypeptide comprising the mature form of an amino acid sequenced selected from the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 226.
2. An isolated polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 226.
3. An isolated polypeptide comprising an amino acid sequence which is at least 95% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 226.
4. An isolated polypeptide, wherein the polypeptide comprises an amino acid sequence comprising one or more conservative substitutions in the amino acid sequence selected from the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 226.
5. The polypeptide of claim 1 wherein said polypeptide is naturally occurring.
6. A composition comprising the polypeptide of claim 1 and a carrier.
7. A kit comprising, in one or more containers, the composition of claim 6.
8. The use of a therapeutic in the manufacture of a medicament for treating a syndrome associated with a human disease, the disease selected from a pathology associated with the polypeptide of claim 1, wherein the therapeutic comprises the polypeptide of claim 1.

9. A method for determining the presence or amount of the polypeptide of claim 1 in a sample, the method comprising:

- (a) providing said sample;
  - (b) introducing said sample to an antibody that binds immunospecifically to the polypeptide; and
  - (c) determining the presence or amount of antibody bound to said polypeptide,
- thereby determining the presence or amount of polypeptide in said sample.

10. A method for determining the presence of or predisposition to a disease associated with altered levels of expression of the polypeptide of claim 1 in a first mammalian subject, the method comprising:

- a) measuring the level of expression of the polypeptide in a sample from the first mammalian subject; and
- b) comparing the expression of said polypeptide in the sample of step (a) to the expression of the polypeptide present in a control sample from a second mammalian subject known not to have, or not to be predisposed to, said disease,

wherein an alteration in the level of expression of the polypeptide in the first subject as compared to the control sample indicates the presence of or predisposition to said disease.

11. A method of identifying an agent that binds to the polypeptide of claim 1, the method comprising:

- (a) introducing said polypeptide to said agent; and
- (b) determining whether said agent binds to said polypeptide.

12. The method of claim 11 wherein the agent is a cellular receptor or a downstream effector.

13. A method for identifying a potential therapeutic agent for use in treatment of a pathology, wherein the pathology is related to aberrant expression or aberrant physiological interactions of the polypeptide of claim 1, the method comprising:

- (a) providing a cell expressing the polypeptide of claim 1 and having a property or function ascribable to the polypeptide;
- (b) contacting the cell with a composition comprising a candidate substance; and
- (c) determining whether the substance alters the property or function ascribable to the polypeptide;

whereby, if an alteration observed in the presence of the substance is not observed when the cell is contacted with a composition in the absence of the substance, the substance is identified as a potential therapeutic agent.

14. A method for screening for a modulator of activity of or of latency or predisposition to a pathology associated with the polypeptide of claim 1, said method comprising:

- (a) administering a test compound to a test animal at increased risk for a pathology associated with the polypeptide of claim 1, wherein said test animal recombinantly expresses the polypeptide of claim 1;
- (b) measuring the activity of said polypeptide in said test animal after administering the compound of step (a); and
- (c) comparing the activity of said polypeptide in said test animal with the activity of said polypeptide in a control animal not administered said polypeptide, wherein a change in the activity of said polypeptide in said test animal relative to said control animal indicates the test compound is a modulator activity of or latency or predisposition to, a pathology associated with the polypeptide of claim 1.

15. The method of claim 14, wherein said test animal is a recombinant test animal that expresses a test protein transgene or expresses said transgene under the control of a promoter at an increased level relative to a wild-type test animal, and wherein said promoter is not the native gene promoter of said transgene.

16. A method for modulating the activity of the polypeptide of claim 1, the method comprising contacting a cell sample expressing the polypeptide of claim 1

with a compound that binds to said polypeptide in an amount sufficient to modulate the activity of the polypeptide.

17. A method of treating or preventing a pathology associated with the polypeptide of claim 1, the method comprising administering the polypeptide of claim 1 to a subject in which such treatment or prevention is desired in an amount sufficient to treat or prevent the pathology in the subject.

18. The method of claim 17, wherein the subject is a human.

19. A method of treating a pathological state in a mammal, the method comprising administering to the mammal a polypeptide in an amount that is sufficient to alleviate the pathological state, wherein the polypeptide is a polypeptide having an amino acid sequence at least 95% identical to a polypeptide comprising the amino acid sequence selected from the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 226 or a biologically active fragment thereof.

20. An isolated nucleic acid molecule comprising a nucleic acid sequence selected from the group consisting of SEQ ID NO:2n-1, wherein n is an integer between 1 and 226.

21. The nucleic acid molecule of claim 20, wherein the nucleic acid molecule is naturally occurring.

22. A nucleic acid molecule, wherein the nucleic acid molecule differs by a single nucleotide from a nucleic acid sequence selected from the group consisting of SEQ ID NO: 2n-1, wherein n is an integer between 1 and 226.

23. An isolated nucleic acid molecule encoding the mature form of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 226.

24. An isolated nucleic acid molecule comprising a nucleic acid selected from the group consisting of 2n-1, wherein n is an integer between 1 and 226.

25. The nucleic acid molecule of claim 20, wherein said nucleic acid molecule hybridizes under stringent conditions to the nucleotide sequence selected from the group consisting of SEQ ID NO: 2n-1, wherein n is an integer between 1 and 226, or a complement of said nucleotide sequence.
26. A vector comprising the nucleic acid molecule of claim 20.
27. The vector of claim 26, further comprising a promoter operably linked to said nucleic acid molecule.
28. A cell comprising the vector of claim 26.
29. An antibody that immunospecifically binds to the polypeptide of claim 1.
30. The antibody of claim 29, wherein the antibody is a monoclonal antibody.
31. The antibody of claim 29, wherein the antibody is a humanized antibody.
32. A method for determining the presence or amount of the nucleic acid molecule of claim 20 in a sample, the method comprising:
- (a) providing said sample;
  - (b) introducing said sample to a probe that binds to said nucleic acid molecule; and
  - (c) determining the presence or amount of said probe bound to said nucleic acid molecule,
- thereby determining the presence or amount of the nucleic acid molecule in said sample.
33. The method of claim 32 wherein presence or amount of the nucleic acid molecule is used as a marker for cell or tissue type.

34. The method of claim 33 wherein the cell or tissue type is cancerous.

35. A method for determining the presence of or predisposition to a disease associated with altered levels of expression of the nucleic acid molecule of claim 20 in a first mammalian subject, the method comprising:

- a) measuring the level of expression of the nucleic acid in a sample from the first mammalian subject; and
- b) comparing the level of expression of said nucleic acid in the sample of step (a) to the level of expression of the nucleic acid present in a control sample from a second mammalian subject known not to have or not be predisposed to, the disease;

wherein an alteration in the level of expression of the nucleic acid in the first subject as compared to the control sample indicates the presence of or predisposition to the disease.

36. A method of producing the polypeptide of claim 1, the method comprising culturing a cell under conditions that lead to expression of the polypeptide, wherein said cell comprises a vector comprising an isolated nucleic acid molecule comprising a nucleic acid sequence selected from the group consisting of SEQ ID NO:2n-1, wherein n is an integer between 1 and 226.

37. The method of claim 36 wherein the cell is a bacterial cell.

38. The method of claim 36 wherein the cell is an insect cell.

39. The method of claim 36 wherein the cell is a yeast cell.

40. The method of claim 36 wherein the cell is a mammalian cell.

41. A method of producing the polypeptide of claim 2, the method comprising culturing a cell under conditions that lead to expression of the polypeptide, wherein said cell comprises a vector comprising an isolated nucleic acid

molecule comprising a nucleic acid sequence selected from the group consisting of SEQ ID NO:2n-1, wherein n is an integer between 1 and 226.

42. The method of claim 41 wherein the cell is a bacterial cell.
43. The method of claim 41 wherein the cell is an insect cell.
44. The method of claim 41 wherein the cell is a yeast cell.
45. The method of claim 41 wherein the cell is a mammalian cell.